Friedrich-Schiller-Universität Jena
Wirtschaftswissenschaftliche Fakultät
Lehrstuhl Wirtschaftsinformatik
Prof. Dr. J. Ruhland

Fake News – Wirkmechanismen, Verbreitung und Erkennung insbesondere in Social Media

Masterarbeit

Zur Erlangung des Grades Master of Science (M.Sc.)

im Studiengang Wirtschaftsinformatik

Eingereicht von:
Paulina Schindler

Gutachter:
Universitätsprofessor
Dr. Johannes Ruhland

Jena, 23.11.20

Inhaltsverzeichnis

I Abbildungsverzeichnis

II Tabellenverzeichnis

III Abkürzungsverzeichnis

Captcha	Completely automated public Turing test to tell computers and humans apart
URL	Uniform Resource Locator, Adresse einer Webseite im Internet

1 Einleitung

Der Begriff ‚Fake News' hat mittlerweile Bekanntheit erreicht. So wurde er zum Anglizismus des Jahres 2016 gewählt.[1] In einer Umfrage der PricewaterhouseCoopers GmbH Wirtschaftsprüfungsgesellschaft (pwc) aus dem Jahr 2019 gaben 31 % der befragten Deutschen an, bereits Fake News wahrgenommen zu haben, 28 % antworteten mit ‚weiß nicht'.[2] Die Aufklärung über Fake News hat aber noch nicht jeden möglichen Empfänger[3] erreicht. 44 % aller Befragten fühlten sich ‚eher schlecht' oder ‚gar nicht' über Fake News aufgeklärt.[4] Ebenso erwarteten 44 %, mit ‚sehr hoher' oder ‚eher hoher' Wahrscheinlichkeit Fake News nicht zu erkennen und dadurch Beeinflussung bezüglich der Europawahl zu erfahren.[5] In Zusammenhang mit der 2020 weltweiten COVID-19-Pandemie wurde im ‚American Journal of Tropical Medicine and Hygiene' eine Studie veröffentlicht, in der die Folgen von Gerüchten, Stigmata und Verschwörungstheorien mittels der Überwachung von Nachrichten auf Onlineplattformen untersucht wurden.[6] So wurden fehlerhafte, verbreitete Informationen mit Gesundheitsschäden und Todesfällen in Verbindung gebracht, z. B. bei angeblichen Heilmitteln. Die Gesundheit der Bevölkerung sei laut der Studie potenziell gefährdet. Mit dem Einsatz neuer Medien, etwa des Internets, kann die Verbreitung von Fake News unterstützt werden. Fake News zeigen sich somit als potenziell gefährliches Instrument, das Auswirkungen auf die Öffentlichkeit haben kann und aktuell Relevanz besitzt. Dadurch sind in den letzten Jahren zahlreiche Forschungsvorhaben entstanden, die sich dem Thema ‚Fake News' annehmen, vor allem dem Auftreten von Fake News auf Social Media und anderen Onlineplattformen. Social Media werden definiert als die Menge der digitalen Plattformen im Internet, die für die Allgemeinheit zugänglich sind.[7] Dabei bieten sie Nutzern die Möglichkeit zur Kommunikation und zum Austausch von Inhalten. Es haben sich zahlreiche Betrachtungsweisen von Fake News und Ansätze hinsichtlich der Verbreitung und Erkennung herausgebildet. Um gegen die Verbreitung von Fake News vorgehen zu können, ist ein umfassendes Verständnis der zugehörigen Mechanismen und Verbreitungsmethoden von Bedeutung. Dies kann eine

[1] Vgl. Stefanowitsch, A. (2017): Anglizismus des Jahres 2016
[2] Vgl. PricewaterhouseCoopers (2019): Fake News, S.9
[3] Im Rahmen dieser Arbeit wird aufgrund der besseren Lesbarkeit stets die männliche Form verwendet, aber alle Geschlechter sollen damit angesprochen werden.
[4] Vgl. PricewaterhouseCoopers (2019): Fake News, S.5
[5] Vgl. PricewaterhouseCoopers (2019): Fake News, S.12
[6] Vgl. Islam et al. (2020): COVID-19-Related Infodemic, S. 4ff.
[7] Vgl. Burkart, R. (2019): Kommunikationswissenschaft, S.37f.

anschließende automatisierte Erkennung erleichtern. Diese Aspekte werden im Rahmen dieser Arbeit näher betrachtet.

Ziel dieser Arbeit ist es, den aktuellen Stand der Forschung zum Thema ‚Fake News' insbesondere im Bereich Social Media wiederzugeben und eine einheitliche Betrachtung und eine grundlegende Systematisierung der in der Forschung vorhandenen Ansätze zu bieten. Damit soll für zukünftige Forschungs- und Recherchevorhaben die Grundlage für ein Verständnis verschiedener Phänomene im Bereich Fake News gelegt und die Kommunikation untereinander erleichtert werden. Dafür wird eine Literaturrecherche (in einer Kombination aus systematischer Recherche und Schneeballsystem) in Bibliotheksdatenbanken und Google Scholar durchgeführt, um die vorhandene Literatur zu Fake News zu untersuchen und die daraus gewonnenen Erkenntnisse adäquat einzuordnen. Die Literaturrecherche eignet sich hierbei besonders, da in dieser Arbeit ein erster Vorstoß zur Systematisierung in diesem Bereich angestrebt wird.

Im Rahmen dieser Arbeit wird dafür zunächst eine einheitliche Verständnisgrundlage geschaffen, indem die verschiedenen Definitionen, mit denen Fake News häufig in der Literatur charakterisiert werden, in Kapitel 2 vorgestellt werden. Teilweise werden dafür Systeme zur Unterscheidung verschiedener Arten fehlerhafter oder schädlicher Informationen eingesetzt, um Sachverhalte gezielter bezeichnen zu können. Um zu verstehen, was Fake News in der Gesellschaft und bei Individuen wirksam macht, werden in Kapitel 3 im Anschluss Wirkmechanismen im Zusammenhang mit Fake News betrachtet. Da Fake News erst wirken können, wenn sie verbreitet werden und ihr beabsichtigtes Publikum erreichen, werden im folgenden Kapitel 4 die generell möglichen Verbreitungswege von Fake News detailliert aufgeschlüsselt. Dabei wird außerdem auf die Relevanz von Social Media als möglicher Verbreitungsweg eingegangen. Im darauffolgenden Kapitel 5 wird untersucht, wie die Verbreitung von Fake News in Social Media im Speziellen durchgeführt wird. Als Maßnahme gegen diese Verbreitungstechniken wurden in der Forschung bisher einige Möglichkeiten zur automatisierten Erkennung von Fake News erarbeitet, die auf Social-Media-Plattformen angewandt werden können. Diese Maßnahmen werden in Kapitel 6 systematisiert, um die grundlegenden Ansätze gezielt hervorheben zu können. Abschließend werden in Kapitel 7 zunächst eventuell unerforschte Bereiche, Problematiken oder Wissenslücken zu Fake News in Social Media dargestellt. Darauf folgt mit Kapitel 8 eine kritische Würdigung der vorliegenden Arbeit, in der die Limitationen aufgeführt werden. Zuletzt findet sich das Fazit.

2 Was sind Fake News?

Der Begriff ‚Fake News' erfuhr im Jahr 2016 zunehmende Verbreitung, wurde aber in den USA bereits ab dem späten 19. Jahrhundert in Lexika eingetragen. Zuvor wurde der Begriff ‚False News' verwendet.[8] Während der Begriff ‚Fake News' ursprünglich genutzt wurde, um erfundene oder falsche Nachrichten zu bezeichnen[9] oder für bestimmte Inhaltsformen wie Parodien, politische Satire und Propaganda eingesetzt wurde, wird er heute eher verwendet, um Falschnachrichten in Social Media zu bezeichnen oder kritische Arbeiten von Nachrichtenagenturen zu untergraben.[10]

Der Begriff ‚Fake News' ist uneinheitlich definiert. In verschiedenen Quellen werden Fake News unterschiedliche Merkmale zugeschrieben, wobei es zu Überschneidungen, teilweise aber auch zu Widersprüchen kommen kann. Eine direkte Übersetzung von ‚Fake' und ‚News' führt zur Bedeutung von falschen oder gefälschten Nachrichten. Menschen, die Fake News verfassen bzw. erstellen, werden im Rahmen dieser Arbeit als ‚Ersteller' bezeichnet, Menschen, die Fake News im Anschluss konsumieren, als ‚Empfänger'.

Zahlreiche Definitionen haben gemeinsam, dass ‚Fake News' als absichtlich und bewusst täuschend beschrieben werden.[11] Andere Quellen räumen die Möglichkeit ein, dass Fake News auch unbeabsichtigt, z. B. durch Fehler oder Nachlässigkeit, entstehen können.[12] Abgesehen von der reinen Täuschungsabsicht werden dem Fake-News-Ersteller auch andere Motivationen, z. B. politische Ideologien oder finanzielle Ziele, unterstellt.[13] Ebenso definieren einige Quellen Fake News als Medien, die in einer Art und Weise geschrieben sind, dass sie Nachrichtencharakter aufweisen.[14] Für einige Autoren ist die Verbreitung online ein bedeutender Aspekt von Fake News[15] oder sogar charakterisierend[16], während

[8] Vgl. McManus, C. und Michaud, C. (2018): Never Mind the Buzzwords, S. 15
[9] Vgl. Borchers, C. (2017): Fake News Has Now Lost All Meaning
[10] Vgl. Tandoc, E. et al. (2018): Defining „Fake News", S.138
[11] Vgl. Allcott, H. und Gentzkow, M. (2017): Fake News in the 2016 Election, S.214; vgl. Sullivan, M. (2017): Time to retire term „fake news"
[12] Vgl. McManus, C. und Michaud, C. (2018): Never Mind the Buzzwords, S. 19
[13] Vgl. Allcott, H. und Gentzkow, M. (2017): Fake News in the 2016 Election, S.217
[14] Vgl. Tandoc, E. et al. (2018): Defining „Fake News", S.138f.; vgl. Horne, B. und Adali, S. (2017): This Just In, S.1, 7
[15] Vgl. Klein, D. und Wueller, J. (2017): Fake News: A Legal Perspective, S.1, 6
[16] Vgl. Bounegru, L. et al. (2018): A Field Guide To „Fake News", S.8

in anderen Quellen darauf kein spezielles Augenmerk gelegt wird.[12] Mit einigen Definitionen wird gefordert, dass Fake News vollständig falsch sein müssen, d. h., dass Fakten nicht deren Basis bilden dürfen[17], was die Frage der Einordnung von Halbwahrheiten und der Manipulation des Kontexts mit einem ‚wahren Kern' aufwirft. Tandoc et al. begegnen diesem Problem, indem sie hohe und niedrige Level von ‚Faktizität' unterscheiden.[18] Ein anderer Ansatz der Argumentation ist es, etwas lediglich dann als ‚Fake News' zu bezeichnen, wenn die angestrebte Täuschung auch gelungen ist, da es sich ansonsten um Fiktion handele.[18] Andere Autoren wiederum sind der Meinung, dass Fake News nicht zwingend geglaubt werden müssen, um als solche zu gelten.[19] Im Gegensatz zu Lügen haben Fake News weniger sozial motivierte Zwecke, etwa sich selbst zu schützen oder Schaden zu vermeiden, sondern dienen jenen, die sie erstellen, zur finanziellen oder politischen Zielerreichung oder der Eigenförderung.[20] Zusätzlich werden Fake News häufig in großem Umfang verbreitet und sind nicht nur auf einzelne Individuen ausgerichtet. Eine linguistische Ähnlichkeit ist dennoch möglich.

Es zeigt sich, dass der Begriff ‚Fake News' keine klare Definition aufweist und eine große Bandbreite an Einsatzszenarien denkbar ist. Daher gibt es zunehmend Bestrebungen, den Begriff mit detaillierteren, genau abgegrenzten Unterkategorien zu versehen oder ihn vollständig durch andere Begrifflichkeiten zu ersetzen, um das jeweils vorliegende Phänomen hinreichend beschreiben zu können.

Laut einer Untersuchung von 34 akademischen Artikeln, die zwischen 2003 und 2017 veröffentlicht wurden, handelt es sich bei ‚Fake News' um keinen klar definierten Begriff. Vielmehr beinhaltet er eine große Bandbreite an verschiedenen Auslegungen in unterschiedlichen Kontexten. In wissenschaftlichen Arbeiten, in denen die Bezeichnung ‚Fake News' verwendet wird, konnten von den Autoren der Untersuchung sechs häufig verwendete Unterarten von Fake News identifiziert werden: Satire, Parodie, Erfindung (Fabrication), Manipulation, Propaganda und Werbung.[21] Eine Untersuchung des Parlaments des Vereinigten Königreichs zu ‚Fake News' rät aufgrund der unscharfen Bedeutung von der

[17] Vgl. Paskin, D. (2018): Real or Fake News: Who Knows, S.254; vgl. Allcott, H. und Gentzkow, M. (2017): Fake News in the 2016 Election, S.4
[18] Vgl. Tandoc, E. et al. (2018): Defining „Fake News", S.148
[19] Vgl. Fallis, D. (2015): What Is Disinformation, S.406
[20] Vgl. Pérez-Rosas, V. et al. (2018): Automatic Detection of Fake News, S.3392
[21] Vgl. Tandoc, E. et al. (2018): Defining „Fake News", S.141ff.

Verwendung dieses übergeordneten Begriffs ab.[22] Stattdessen wird die Verwendung von ‚Misinformation', ‚Disinformation' und ‚Malinformation' nach dem ‚Information-Disorder'-Framework von Wardle und Derakshan vorgeschlagen.[23] Misinformation bezeichnet dabei das Verbreiten von falscher Information, ohne dass dabei Schaden zugefügt werden soll. Disinformation ist falsche Information, die bewusst verbreitet wird, um Schaden zu verursachen. Malinformation bezeichnet korrekte Informationen, die verbreitet werden, um Schaden zuzufügen, z. B. das Publikmachen privater Daten. Das Konzept von Mis- und Disinformation wurde bereits vor dem Entstehen des Frameworks verwendet.[24] Fallis beschreibt, dass der Übergang von Mis- zu Disinformation und umgekehrt fließend sein kann, indem z. B. zunächst satirische Misinformation anschließend mit einer Täuschungsabsicht verteidigt oder ursprüngliche Disinformation unwissentlich weiterverbreitet wird.[25]

‚Fake News' können sich sowohl als Mis- als auch als Disinformation zeigen.[26] Da es sich bei Malinformation um korrekte Informationen handelt, statt um im Zusammenhang mit ‚Fake News' zu erwartende tatsächlich gefälschte Informationen, nimmt der Begriff in Bezug auf Fake News eine Sonderrolle ein. Im Rahmen dieser Arbeit soll der Fokus auf Falschinformationen liegen, sodass Malinformation an dieser Stelle nicht näher betrachtet wird.

Während Wardle und Derakshan Mis- und Disinformation auf eine Stufe stellen[27], betrachten andere Autoren Disinformation als eine Unterart von Misinformation.[19] Misinformation und Disinformation voneinander zu unterscheiden kann eine Herausforderung sein, da die Intention bei der Veröffentlichung von Medien häufig unklar ist.[28] Aus diesem Grund werden die beiden Begriffe mitunter austauschbar verwendet. Übergreifend über die Kategorien Mis- und Disinformation unterscheidet Wardle sieben Arten.[29] Diese werden in Tabelle 1 zusammen mit den von Wardle vermuteten Ursachen bzw. Intentio-

[22] Vgl. UK Parliament (2018): Disinformation and „fake news"
[23] Vgl. Wardle, C. und Derakhshan, H. (2017): Information Disorder, S.20
[24] Vgl. Hernon, P. (1995): Disinformation and misinformation through the Internet, S.134
[25] Vgl. Fallis, D. (2015): What Is Disinformation, S.406, 414f.
[26] Vgl. Wardle, C. und Derakhshan, H. (2017): Information Disorder, S.14
[27] Vgl. Wardle, C. und Derakhshan, H. (2017): Information Disorder, S.5
[28] Vgl. Jack, C. (2017): Lexicon of Lies, S.2
[29] Vgl. Wardle, C. (2017): Fake News. It's complicated.

nen dahinter aufgeschlüsselt. Horizontal sind die sieben Arten von Mis- und Disinformation zu sehen, vertikal sind acht Motive für das Erstellen von Fake News notiert. Kreuze in der Tabelle kennzeichnen, dass eine Kombination laut Wardle zutrifft.

	Satire oder Parodie	Falsche Verbindung	Irreführende Inhalte	Falscher Kontext	Betrügerische Inhalte	Manipulierte Inhalte	Erfundene Inhalte
Schlechter Journalismus		X	X	X			
Parodieren	X				X		X
Provozieren					X	X	X
Passion				X			
Parteilichkeit			X	X			
Profit		X			X		X
Politischer Einfluss			X	X		X	X
Propaganda			X	X	X	X	X

Tabelle 1: Arten von Mis- und Disinformation und Motive nach Wardle (2017)[29]

Es zeigt sich, dass die von Wardle vorgeschlagenen Unterarten von Misinformation in einigen Punkten den sechs Unterarten von ‚Fake News' aus Tandoc et al. ähneln, Wardle die einzelnen Aspekte allerdings detaillierter aufschlüsselt. Der Aspekt der Satire und Parodie wird sowohl von Tandoc et al. als auch von Wardle aufgegriffen, ebenso wie das Erfinden und Manipulieren von Inhalten. Propaganda greift Wardle nicht als Unterart von Mis- bzw. Disinformation auf, sondern als Motiv. Werbung ist hingegen bei Wardle keine eigene Kategorie. Es ist davon auszugehen, dass Werbung in allen Unterkategorien vorkommen könnte. Zusätzlich werden noch die falsche Verbindung, irreführende Inhalte, ein falscher Kontext und betrügerische Inhalte als Unterarten eingeführt und somit eine höhere Nuanciertheit erreicht.

Im Rahmen dieser Arbeit wird im Folgenden der Einfachheit halber weiterhin von ‚Fake News' gesprochen. Gemeint ist damit die vom Ersteller von Fake News bewusst ausgelöste Verbreitung von fehlerhafter Information mit Täuschungsabsicht. Da diese im Anschluss von Unwissenden weiterverbreitet werden kann, umschließt dies sowohl Mis- als auch Disinformation. Mis- und Disinformation stehen miteinander in enger Verbindung. Der Übergang ist fließend, da Disinformation, die von einem Nutzer unwissentlich geteilt wird, auch zu Misinformation werden kann.[25] Daher können beide Kategorien schwierig

voneinander getrennt werden. Unabhängig davon, ob (unter Einbezug von Misinformation) damit vom Weiterverbreitenden nachhaltig Schaden verursacht werden soll oder nicht, wird mit dem Verbreiten von falschen Informationen eine Täuschung des Empfängers ausgelöst. Die Wirkmechanismen, die eine Täuschung ermöglichen, begünstigen oder generell mit Fake News einhergehen, werden im folgenden Kapitel besprochen.

3 Wirkmechanismen im Kontext von Fake News

Um zu erfahren, was Fake News in der Gesellschaft und bei Individuen wirksam macht, ist eine Betrachtung der Wirkmechanismen, die im Zusammenhang mit jenen wirken, von Bedeutung. Dabei variiert die Stärke der Wirkung der Mechanismen abhängig von den Eigenschaften des Empfängers.

3.1 Die Wirkmechanismen von Fake News

Die Ersteller von Fake News setzen häufig verschiedene Mechanismen ein, die im Zusammenhang mit Fake News wirken und die Falschinformationen beim Empfänger wirksamer machen können, die Wirkung verstärken oder gegen Gegenargumente immunisieren. Diese Mechanismen können von den Erstellern von Fake News genutzt werden oder wirken zusammen mit Fake News unbewusst. Ebenso werden manche Wirkmechanismen, die für die Wirksamkeit von Fake News eine bedeutende Rolle spielen können, nicht zwingend zielgerichtet vom Ersteller der Fake News eingesetzt. Vielmehr können auch Mechanismen eine Rolle spielen, die sich durch das Umfeld bzw. die Umwelt des Empfängers oder den Umgang mit Fake News ergeben. In Tabelle 2 werden verbreitete Wirkmechanismen, die in Zusammenhang mit Fake News Einfluss nehmen können, alphabetisch sortiert, nach ihrem gebräuchlichsten Namen (wenn vorhanden) aufgelistet und erklärt.

Wirkmechanismus	Erläuterung
Astroturfing	Beim Astroturfing wird versucht, einen inkorrekten Eindruck von der öffentlichen Meinung zu vermitteln, z. B. indem eine große Mehrheit für eine bestimmte Entscheidung vorgetäuscht wird. Im Gegensatz zu einer ‚Graswurzelbewegung' steht dahinter aber nicht tatsächlich die Bevölkerung, sondern das Astroturfing wird von einem verdeckten Initiator organisiert.[30]

[30] Vgl. Voss, K. (2010): Grassrootscampaigning und Chancen durch neue Medien

Auflösung von Gerüchten	Gerüchte, die inkorrekt sind, benötigen in Social Media länger, um aufgelöst zu werden als wahre Gerüchte.[31] Nicht verifizierte Gerüchte werden häufig früher geteilt und erreichen eine größere Nutzerbasis als aufgelöste Gerüchte.[31]
Availability-Cascade	Individuen tendieren dazu, die Ansichten anderer anzunehmen, wenn diese Ansichten in ihrem sozialen Umfeld an Popularität gewinnen.[32]
Availability-Heuristic (Verfügbarkeitsheuristik)	Die Wahrscheinlichkeit von Ereignissen wird danach bemessen, wie verfügbar ein ähnliches Ereignis in der Erinnerung ist. Eine kürzliche oder häufige Berichterstattung zu bestimmten Ereignissen sorgt also dafür, dass diese für wahrscheinlicher gehalten werden.[33]
Backfire-Effect	Bei der Bereitstellung von Gegenargumenten konnte festgestellt werden, dass Probanden nach der Korrektur einer (politischen) Information noch stärker an die ursprüngliche, falsche Information glaubten.[34] Es wird vermutet, dass dieser Effekt lediglich in speziellen Situationen auftritt, da er mit einem anderen Versuchsaufbau nicht nachgewiesen werden konnte.[35]
Bandwagon-Effect (Mitläufer-Effekt)	Dies bezeichnet die Annahme, dass, wenn andere Personen etwas als gut empfinden, es auch von der eigenen Person gut zu beurteilen sein wird.[36] Es wird sich in der Meinungsbildung an anderen Personen orientiert. Dieses Phänomen konnte beispielsweise auch bei Online-Reviews festgestellt werden.[37]
Clickbait	Wissenslücken (Information-Gap), die durch Titel von Nachrichten aufgebaut werden, wecken beim potenziellen Leser Neugier für den restlichen Artikel. Oftmals wird dabei eine Vorwärtsreferenz (forward reference) verwendet, die auf weitere Informationen im Artikel verweist.[38]

[31] Vgl. Zubiaga, A. et al. (2016): Rumors in Social Media, S.26
[32] Vgl. Kuran, T. und Sunstein, C. (1999): Availability Cascades and Risk Regulation, S.683
[33] Vgl. Tversky, A. und Kahneman, D. (1973): Availability, S.228f.
[34] Vgl. Nyhan, B. und Reifler, J. (2010): When Corrections Fail, S.323
[35] Vgl. Swire, B. et al. (2017): Processing political misinformation, S.17
[36] Vgl. Sundar, S. et al. (2008): The Bandwagon Effect, S.3454
[37] Vgl. Sundar, S. et al. (2008): The Bandwagon Effect, S.3457f.
[38] Vgl. Blom, J. und Hansen, K. (2015): Click bait, S.87f.

Confirmation-Bias (Bestätigungsfehler)	Menschen bevorzugen unbewusst Informationen, die sich mit ihrer eigenen Meinung decken und halten diese für glaubwürdiger.[39] Es wird vermutet, dass dies zur Entstehung von Echo Chambers und Filterblasen beiträgt.[40]
Conservatism-Bias	Dies bezeichnet die Tendenz von Individuen, ihre Einstellung bei neuer Information unzureichend anzupassen.[41] Glaubt eine Person also bereits an Fake News, ist die Information schwierig zu korrigieren.
Continued-Influence-Effect	Auch die Negation und die Korrektur inkorrekter ursprünglicher Informationen können deren Wirkung meist nicht vollständig rückgängig machen und beeinflussen den Empfänger weiter.[42] Dieser Effekt wird abgeschwächt, wenn statt einer einfachen Korrektur eine passende alternative Erklärung für ein Szenario angeboten wird.[43] Dies wird auch als ‚Belief-Perseverance' bezeichnet.[44]
Echo-Chamber-Effect (Echokammereffekt)	Wenn sich Nutzer hauptsächlich mit anderen Nutzern bzw. Einrichtungen auseinandersetzen, die eine ähnliche Meinung vertreten wie sie selbst, entsteht eine Echokammer. Die Nutzer bestärken sich somit gegenseitig in ihrer Meinung.[45] Der Confirmation-Bias funktioniert auf eine ähnliche Weise. Oftmals werden Nutzer aber nicht vollständig isoliert, sondern bleiben weiterhin mit gegensätzlichen Inhalten konfrontiert.[46]
Emotional-Memory-Enhancement	Emotional aufgeladene Informationen werden besser behalten als neutrale Informationen.[47] Suggestion wirkt hierbei noch stärker als reine Emotionalität.[48]

[39] Vgl. Nickerson, R. (1998): Confirmation Bias, S.175f.
[40] Vgl. Lazer, D. et al. (2017): Combating Fake News, S.5
[41] Vgl. Ward, E. (1982): Conservatism in Human Information Processing, S.359
[42] Vgl. Ross, L. et al. (1975): Perseverance in Self-Perception, S.888
[43] Vgl. Johnson, H. und Seifert, C. (1994): Sources of the Continued Influence Effect, S.1431
[44] Vgl. Cobb, M. et al. (2013): Beliefs Don't Always Persevere, S.307
[45] Vgl. Colleoni, E. et al. (2014): Echo Chamber or Public Sphere, S.317
[46] Vgl. Zuiderveen Borgesius, F. et al. (2016): Should We Worry about Filter Bubbles, S.6f.
[47] Vgl. Van Damme, I. und Smets, K. (2014): The power of emotion, S. 310
[48] Vgl. Van Damme, I. und Smets, K. (2014): The power of emotion, S. 315

Filter Bubble (Filterblase)	Dieser Begriff bezeichnet Informationsblasen, die insbesondere in Social Media entstehen und in denen Algorithmen Inhalte auswählen bzw. vorfiltern, die dem Nutzer anschließend angezeigt werden. Diese Inhalte entsprechen häufig den bereits bestehenden Interessen. Nutzer sind sich oft der Filterblase nicht bewusst.[49] Somit werden keine gegenteiligen Meinungen angezeigt, die Fake News entkräften könnten.
Framing-Effect	Kleine Änderungen im Kontext oder der Art und Weise der Informationsübermittlung können zu einer starken Veränderung des Entscheidungsverhaltens führen.[50]
Google-Effect	Menschen tendieren dazu, sich nicht Informationen an sich zu merken, sondern nur, wo diese im Bedarfsfall zu finden sind.[51] Somit könnte nicht vorhandenes Hintergrundwissen nicht gegen Fake News wirken.
Hostile-Media-Effect	Voreingenommene Probanden fühlen sich von der Berichterstattung in den Medien benachteiligt, auch wenn ein Großteil der Rezipienten sie als angemessen wahrnimmt.[52] Dies kann den Glauben an die Korrektur von Fake News durch große Nachrichtenagenturen vermindern.
Illusory-Truth-Effect (Wahrheitseffekt)	Aussagen, die mehrfach gehört werden, wird ein höherer Wahrheitswert zugesprochen als Aussagen, die zum ersten Mal gehört werden. Das heißt, Wiederholung erhöht die Wahrscheinlichkeit, dass eine Aussage als wahr erachtet wird. Dies gilt auch bei einer geringen Plausibilität der Aussage[53] oder im Falle von Warnungen davor.[54] Dieser Effekt wird auch als ‚Validity-Effect' bezeichnet.[55]

[49] Vgl. Pariser, E. (2012): Filter Bubble, S.18
[50] Vgl. Stocké, V. (2002): Framing und Rationalität, S.10
[51] Vgl. Sparrow, B. et al. (2011): Google Effects on Memory, S.778
[52] Vgl. Vallone, R. et al. (1985): The Hostile Media Phenomenon, S.277
[53] Vgl. Fazio, L. et al. (2019): Repetition increases perceived truth, S. 1705
[54] Vgl. Pennycook, G. und Rand, D. (2020): Who falls for fake news, S.186
[55] Vgl. Boehm, L. (1994): The Validity Effect, S.285

Implied-Truth-Effect	Werden andere Nachrichten als Fake News erkannt bzw. gekennzeichnet, eine aber nicht, so wird diese eher als wahr betrachtet.[56]
Misdirecting	Misdirecting wird benutzt, wenn kontextbezogene Hashtags eingesetzt werden, aber über ein völlig anderes Thema berichtet wird.[57] Dadurch wird vom eigentlichen Thema abgelenkt und tatsächliche Informationen gehen in der Menge der Nachrichten unter.[58]
Misinformation-Effect (Fehlinformationseffekt)	Eine auf ein Ereignis folgende, unwahrheitsgemäße Berichterstattung schadet der korrekten Erinnerung an dieses Ereignis.[59]
Negativity-Bias	Menschen haben die Tendenz, negativen Informationen höheres Gewicht beizumessen als positiven Informationen.[60]
Primacy-Effect & Recency-Effect	Informationen, die ein Empfänger zuerst aufnimmt, prägen diesen stärker als die nachfolgenden (Primacy-Effect).[61] Ebenso bleibt die zuletzt aufgenommene Information länger im Gedächtnis (Recency-Effect).[61]
Reputation-Heuristic	Statt die Inhalte einer Quelle zu prüfen, wird die Quelle selbst auf ihre Glaubwürdigkeit geprüft. Hat die Quelle einen guten Ruf oder gilt sie als glaubwürdig, so werden die Informationen eher geglaubt.[62] Gelingt es Fake-News-Erstellern, eine glaubwürdige Quelle zu imitieren, erhöht sich deren Glaubwürdigkeit.
Smoke-Screening	Smoke-Screening funktioniert wie Misdirecting mit dem Unterschied, dass zu einem Hashtag zumindest ähnliche Inhalte gepostet werden.[57]

[56] Vgl. Pennycook, G. et al. (2020): The Implied Truth Effect, S.11
[57] Vgl. Akademische Gesellschaft (2018): How powerful are Social Bots, S.1
[58] Vgl. Abokhodair, N. et al. (2015): Dissecting a Social Botnet, S.849f.
[59] Vgl. Blank, H. und Launay, C. (2014): Protect against the misinformation effect, S. 78
[60] Vgl. Wojciszke, B. et al. (1993): Effects of Information Content, S.327
[61] Vgl. Krosnick, J. und Alwin, D. (1987): An Evaluation of a Cognitive Theory, S.202
[62] Vgl. Metzger et al. (2010): Credibility Evaluation Online, S.426

Tainted-Truth-Effect	Warnungen vor falscher Information, die fälschlicherweise in Bezug auf wahrheitsgetreue Inhalte herausgegeben werden, schaden der Glaubwürdigkeit der wahrheitsgemäßen Informationen.[63]
Third-Person-Effect	Menschen tendieren dazu, zu glauben, dass Massenmedien andere Menschen stärker beeinflussen als sie selbst.[64] Dadurch kann der Einfluss von Fake News auf die eigene Person unterschätzt werden.

Tabelle 2: Übersicht über die Wirkmechanismen von Fake News

Diese Wirkmechanismen können im Zusammenhang mit Fake News relevant werden. Bestimmte Eigenschaften oder Stimmungen, die eine Person auszeichnen, können dazu führen, dass die Empfänger unterschiedlich stark an Fake News glauben oder auch für die Wirkmechanismen mehr oder weniger anfällig sind.

3.2 Personeneigenschaften als moderierende Variable

Die menschliche Erinnerung ist kein unfehlbares Aufnahmegerät, sondern es erfolgt ein Rekonstruktionsprozess, der internen und externen Einflüssen unterworfen ist.[47] Manche Probanden zeigten sich dabei in Studien für externe Einflüsse wie Fake News und deren zugehörige Wirkmechanismen anfälliger als andere. Die Stärke, mit der Wirkmechanismen eine Person beeinflussen, ist abhängig von den Eigenschaften dieser Person und somit variabel. Währen einige Personeneigenschaften die Wirkung der Wirkmechanismen abschwächen können, wird sie durch andere verstärkt.

Probanden, die zu analytischem Denken eher in der Lage waren, hielten in einer Studie von Pennycook und Rand Fake News mit geringerer Wahrscheinlichkeit für wahr.[65] Im Gegensatz dazu scheinen leichtgläubigere Probanden bzw. Probanden, die in Sätze eingebaute, zufällige Buzzword-Wortkombinationen eher als tiefgründig beschrieben, Fake News schlechter erkennen zu können als skeptisch eingestellte.[66] Das bedeutet aber nicht grundsätzlich, dass analytisch denkende Menschen immun gegen die Wirkmechanismen von Fake News sind. So konnte z. B. festgestellt werden, dass der Unterschied zwischen

[63] Vgl. Freeze, M. et al. (2020) Fake Claims of Fake News, S. 24
[64] Vgl. Davidson, W. (1983): The Third-Person Effect in Communication, S.1
[65] Vgl. Pennycook, G. und Rand, D. (2019): Lazy, not biased, S.47
[66] Vgl. Pennycook, G. und Rand, D. (2020): Who falls for fake news, S.196f.

unbekannten und bekannten (d. h. wiederholten – Illusory-Truth-Effect) Schlagzeilen nicht mit dem Maß des analytischen Denken eines Probanden interagiert, sondern beide Aspekte unabhängig wirken. Es wird vermutet, dass dies damit zu begründen ist, dass Wiederholungen auf einer niedrigeren kognitiven Ebene wirken.[66] Dies ist auch für andere Wirkmechanismen von Fake News denkbar. Zum Zeitpunkt der Fake-News-Aufnahme empfundene Emotionen können ebenfalls beeinflussen, wie mit den jeweiligen Informationen umgegangen wird. So konnte bei einer Untersuchung von Weeks aus dem Jahr 2016 die Wirkung von Ärger und Angst auf die unkorrigierte Aufnahme von falschen politischen Informationen untersucht werden. Es zeigte sich, dass verärgerte Probanden die Information eher so interpretierten, dass ihr parteiischer Glaube bestärkt wurde, während Angst die Probanden sich weniger auf ihre Parteizugehörigkeit, sondern auf die Informationen im Artikel stützen ließ.[67] Weiter wird angenommen, dass Personen für Fake News eher anfällig sind, wenn die Informationen die eigene politische Ideologie unterstützen. Dabei ist auch eine Asymmetrie zwischen politischen Lagern (im Experiment Demokraten und Republikaner) bezüglich der Fähigkeit, die Wahrheit von Fake News zu unterscheiden, zu vermuten.[65] Allcott und Gentzkow stellten im Jahr 2017 fest, dass Personen, die Medien intensiv nutzen, mit höherer Wahrscheinlichkeit ideologisch ausgerichteten Artikeln glauben.[68] Ebenso glauben Personen mit abgesonderten Netzwerken eher daran. Personen, die bezüglich ihrer Wahlentscheidung zum Zeitpunkt der Untersuchung unentschieden waren, glaubten weniger an ideologisch ausgerichtete Artikel als Wähler, die sich bereits für eine Partei entschieden hatten.[68] Anfällig für Fake News zeigten sich in einer Studie Teilnehmer, die wahnhaft, dogmatisch oder religiöse Fundamentalisten waren.[69] Dieser Umstand könnte in einem Zusammenhang mit reduziertem analytischen Denken stehen.[69] In Verbindung mit einer Echo Chamber wird der Effekt noch deutlicher, umso extremer die Ansichten sind.[70]

Das Alter einer Person kann ebenso eine Rolle hinsichtlich der Wirkung von Fake News spielen. In einer Untersuchung zum Umgang mit Fake News in Social Media während der Wahlkampagne 2016 in den USA zeigte sich, dass Nutzer, die über 65 Jahre alt waren, nahezu siebenmal mehr Fake-News-Artikel teilten als die jüngste Altersgruppe der 18-

[67] Vgl. Weeks, B. (2015): Emotions, Partisanship, and Misperceptions, S.699, 712
[68] Vgl. Allcott, H. und Gentzkow, M. (2017): Fake News in the 2016 Election, S.230
[69] Vgl. Bronstein, M. et al. (2019): Belief in Fake News, S.115
[70] Vgl. Boutyline, A. und Willer, R. (2017): The Social Structure of Political Echo Chambers, S.565f.

bis 29-Jährigen.[71] Dies ist möglicherweise auf die geringere Erfahrung von älteren Personen im Umgang mit dem Internet zurückzuführen.[72] Im Kontrast dazu gaben Befragte über 60 Jahren in einer Bevölkerungsbefragung in Deutschland an, Nachrichten kritisch zu lesen[73] und weniger mit Fake News in Berührung zu kommen als jüngere Befragte.[74] Gleichzeitig gaben 37 % (und somit die größte Gruppe dieser Kategorie) der Befragten über 60 Jahre an, noch keine Aufklärung über Fake News mitbekommen zu haben.[75] Insgesamt wird von den meisten Befragten die je eigene Altersgruppe als weniger von Fake News gefährdet angesehen.[76] Vor allem bei jüngeren Deutschen und EU-Ausländern sehen alle Befragten insgesamt eine größere Gefahr der Beeinflussung durch Fake News.[77] Es ist davon auszugehen, dass zusätzlich bei der Wahrnehmung der Glaubwürdigkeit von Nachrichten die Kultur, aus der der Empfänger kommt, eine Rolle spielt. So zeigten sich bei einer Untersuchung auf Twitter beispielsweise Unterschiede hinsichtlich der Bewertung der Glaubwürdigkeit von Nachrichten zwischen Nutzern aus den USA und aus China.[78] Ebenso konnte eine unterschiedlich kritische Einstellung bei Nutzern aus verschiedenen Ländern gegenüber dem Inhalt von Nachrichten festgestellt werden.[79]

Jede Person kann also aufgrund ihrer spezifischen Gegebenheiten unterschiedlich stark auf Fake News und deren Wirkmechanismen reagieren. Damit eine solche Reaktion stattfindet, müssen Empfänger in Kontakt mit Fake News kommen. Diesbezüglich existieren prinzipiell verschiedene Verbreitungswege.

[71] Vgl. Guess, A. et al. (2019): Less than you think, S.1, 3
[72] Vgl. Schäffer, B. (2007): The Digital Literacy of Seniors, S.38
[73] Vgl. PricewaterhouseCoopers (2019): Fake News, S.26
[74] Vgl. PricewaterhouseCoopers (2019): Fake News, S.10
[75] Vgl. PricewaterhouseCoopers (2019): Fake News, S.23
[76] Vgl. PricewaterhouseCoopers (2019): Fake News, S.15
[77] Vgl. PricewaterhouseCoopers (2019): Fake News, S.14
[78] Vgl. Yang, J. et al. (2013): Microblog Credibility Perceptions, S.584
[79] Vgl. Shariff, S. et al. (2017): On the credibility perception of news on Twitter, S.794

4 Generelle Verbreitungswege von Fake News

Fake News können auf vielfältige Art und Weise verbreitet werden. Erst durch die Verbreitung und das Erreichen eines Publikums können die Wirkmechanismen Einfluss nehmen. Da Fake News in der Regel möglichst viele Empfänger erreichen sollen, eignen sich für eine Verbreitung vor allem Massenmedien. Diese sind dadurch charakterisiert, dass sie zur öffentlichen Verbreitung von Informationen eingesetzt werden und sich prinzipiell in vier Oberkategorien einteilen lassen: Printmedien, audiovisuelle Medien, massenhaft verbreitete Speichermedien und Webseiten im Internet.[80] Diese Kategorien sind in Abbildung 1 aufgeführt. Beispiele für bekannte Ausprägungen befinden sich darunter. Während zu den Printmedien beispielsweise Bücher, Zeitungen, Plakate und Flyer zu zählen sind, beinhalten audiovisuelle Medien unter anderem das Fernsehen, Filme und Radiosender. Zu massenhaft verbreiteten Speichermedien lassen sich großflächig vertriebene CDs oder DVDs zählen. Webseiten im Internet beinhalten Onlineangebote und Vernetzungsmöglichkeiten, z. B. Social Media. Diese vier Kategorien an Massenmedien werden als generelle Verbreitungswege betrachtet.

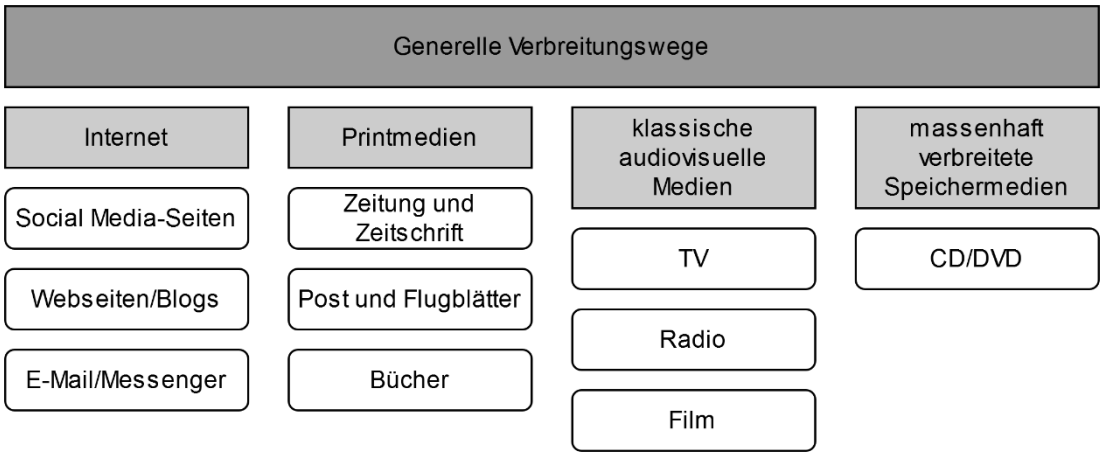

Abbildung 1: Generelle Verbreitungswege von Fake News

Prinzipiell ist in allen Kategorien ein Einsatz von Fake News denkbar, wenn auch der Einsatz in manchen Kategorien aufgrund der Möglichkeiten moderner Technologie (z. B. automatisiert) praktikabler erscheint als in anderen (beispielsweise bei der Nutzung von Speichermedien, deren Erstellung und Verbreitung aufgrund ihres physischen Charakters mit höherem Kosten- und Zeitaufwand zu erwarten ist). In jeder dieser Kategorien gibt

[80] Vgl. Burkart, R. (2019): Kommunikationswissenschaft, S.108f.

es eine Quelle bzw. einen Ersteller von Fake News, der den Empfängern mehr oder weniger offensichtlich (oder z. B. getarnt als eine andere Quelle) gegenübertritt. Dieser Aspekt wird als zwischenmenschliche Quelle bezeichnet, da der Fokus auf die interpersonale Beziehung zwischen dem Ersteller und dem Empfänger der Fake News gerichtet ist. Auch wenn beispielsweise die Automatisierung für die Verbreitung von Fake News genutzt wird, wird diese von einem Menschen im Hintergrund ausgelöst, der in diesem Fall der (womöglich verschleierte) Ersteller ist.

Während Fake News häufig in Textform verbreitet werden, ist auch die Verbreitung als Bild, Ton oder als eine Kombination dieser Medien denkbar. Ähnlich wie tatsächliche Nachrichtenmeldungen manipuliert werden können, können auch Bild oder Ton heutzutage mit technischen Hilfsmitteln verändert oder in einen neuen Kontext gesetzt werden, um Fake News zu verbreiten. Mithilfe von künstlicher Intelligenz können realistisch wirkende, schwer zu erkennende Fälschungen erstellt werden, sogenannte ‚Deepfakes'.[81] Solche Fake News können häufig im Bereich des Internets angetroffen werden. Im Folgenden hat diese Arbeit vor allem den Social-Media-Bereich zum Thema.

[81] Vgl. Kietzmann, J. et al. (2020): Deepfakes, S.135

5 Verbreitung von Fake News in Social Media

Es gibt unterschiedliche Möglichkeiten, Fake News (automatisiert) zu Erstellen und zu Verbreiten. Der Fokus soll hierbei auf den Eigenheiten von Fake News in Social Media liegen. Zu verstehen, wie die Erstellung und Verbreitung von Fake News erreicht werden kann, erleichtert die anschließende Erkennung jener.

5.1 Social Media als Plattform für Fake News

Fake News können prinzipiell in jedem Massenmedium auftreten. Allerdings nimmt das Internet, insbesondere auch der Social-Media-Bereich, eine prominente Stellung ein. Potenzielle Empfänger, die sich auf Social Media befinden, werden im Folgenden als ‚Nutzer' bezeichnet.

Social-Media-Plattformen sind mit einer privaten Nutzung bei durchschnittlich 55 % aller Befragten verbreitet.[82] Vor allem junge Nutzer zwischen 14 und 29 Jahren nutzen vermehrt Social-Media-Plattformen[83] und informieren sich damit über das aktuelle Geschehen.[84] Daher ist davon auszugehen, dass mit dem Älterwerden der jüngeren Generation Social Media weiterhin oder zunehmend genutzt werden. Dies bedeutet aber nicht, dass Fake News eher von jungen Menschen erkannt werden können, wie in einer Studie aus dem Jahr 2016 mit Schülern aus den USA festgestellt wurde.[85] In den vergangenen Jahren nahm die Nutzung von Social Media als Nachrichtenquelle von 18 % im Jahr 2013 auf 37 % im Jahr 2020 zu und übersteigt somit mittlerweile die Nutzung von Printmedien als Nachrichtenquelle im Jahr 2020 mit 33 %.[86] Während 54 bis 59 % der jüngeren Nutzer zwischen 18 und 39 Jahren laut einer Umfrage von PricewaterhouseCoopers aus dem Jahr 2018 zumindest einem sozialen Medium vertrauen, ist dieses Vertrauen bei älteren Befragten geringer.[87] Das Vertrauen in Social Media sank von 2016 bis 2018 laut den Befragten allerdings.[88] Das könnte auch auf Fake News zurückzuführen sein, die 2016 im

[82] Vgl. Statistisches Bundesamt (2020): Personen mit Internetaktivitäten
[83] Vgl. ARD und ZDF (2019): Onlinecommunities
[84] Vgl. PricewaterhouseCoopers (2018): Vertrauen in Medien, S.6
[85] Vgl. Wineburg, S. et al. (2016): Evaluating Information, S.4
[86] Vgl. Newman, N. et al. (2020): Digital News Report 2020, S.71
[87] Vgl. PricewaterhouseCoopers (2018): Vertrauen in Medien, S.14
[88] Vgl. PricewaterhouseCoopers (2018): Vertrauen in Medien, S.15

Rahmen des US-Wahlkampfs häufig thematisiert wurden.[89]

Social Media bieten ein großes Einsatzgebiet für Fake News. Durch die Reichweite einer solchen Plattform bzw. die plattformübergreifende Reichweite ist auch der Aktionsradius von Fake News potenziell hoch. Laut einer Untersuchung aus dem Jahr 2019 werden Fake News in Deutschland auf Facebook sechsmal häufiger geteilt als professionelle Informationen.[90] Da keine physischen Kopien erstellt werden müssen, sind Fake News dort außerdem theoretisch unendlich replizierbar. Ihre Verbreitung erfolgt häufig schneller und durchdringt Netzwerke tiefer als die Verbreitung korrekter Informationen und zieht somit eine schnelle Reaktion seitens der Empfänger nach sich.[91] Die Wirkung von Fake News kann daher bereits eintreten, bevor Gegendarstellungen oder Korrekturen veröffentlicht werden. Filter Bubbles und Echo Chambers, die ebenfalls in Verbindung mit Social Media auftreten können, verstärken diesen Effekt und können den Kontakt eines Nutzers mit Korrekturmaßnahmen verhindern. In einer Umfrage von YouGov aus dem Jahr 2017 wurden Facebook und Twitter als am anfälligsten für die Veröffentlichung und Verbreitung von Fake News seitens der Befragten beurteilt.[92] Allerdings gibt es auch Nutzer, die ihre Fähigkeit, Fake News zu erkennen, überschätzen und somit selbst anfällig sind. Auch wenn Befragte in einer Studie aus Singapur aus dem Jahr 2018 angaben, Fake News sicher erkennen zu können, gelang ihnen das nicht immer. 90 % der Befragten aus Singapur identifizierten mindestens eine von fünf Fake-News-Schlagzeilen als wahr. Es gab keine Korrelation zwischen der Überzeugung, Fake News erkennen zu können, und der tatsächlichen Fähigkeit dazu.[93]

Abgesehen von der Möglichkeit der kostengünstigen, schnellen und großflächigen Verbreitung bieten Social Media in Bezug auf Fake News noch andere Vorteile, z. B. das spezifische Ansprechen eines einzelnen Nutzers mit auf ihn zugeschnittenen Fake News (Microtargeting). Microtargeting ermöglicht es, Nutzer auf Social Media individuell anzusprechen und ihnen die für sie jeweils passendsten Fake News zu präsentieren, deren Glaubwürdigkeit der Empfänger als hoch bewertet. In keinem anderen Massenmedium ist eine dermaßen gezielte Ansprache mit vergleichsweise geringem Aufwand möglich,

[89] Vgl. Allcott, H. und Gentzkow, M. (2017): Fake News in the 2016 Election, S.211ff.
[90] Vgl. Marchal, N. et al. (2019): Junk News During the EU Parliamentary Elections, S.4
[91] Vgl. Vosoughi, S. et al. (2018): The spread of true and false news online, S.1146
[92] Vgl. YouGov (2017): Alles Fake, S.12
[93] Vgl. Huiwen, N. (2018): Most people say they can spot fake news but falter when tested

insbesondere auf Basis der bereits vom Nutzer hinterlegten und zumeist öffentlich zugänglichen Daten. Durch die Ansammlung von Big Data und anschließendes Data-Mining können Nutzer auch in großer Stückzahl automatisiert mit individuell angepassten Fake News versorgt werden. Mithilfe von Big-Data-Analysen können Nutzer nicht nur nach klassischen soziodemographischen Merkmalen (z. B. Alter und Geschlecht) unterteilt werden, sondern auch weitere Informationen erhoben werden, die die Nutzer möglicherweise nicht bewusst preisgeben wollen. In einer Studie aus dem Jahr 2013 wurde gezeigt, dass sich allein durch die Auswertung von Likes eines Nutzers auf Facebook zahlreiche Informationen zu dessen Persönlichkeit und zu sensiblen Inhalten wie der sexuellen Orientierung, der politischen Einstellung, der Religiosität, der ethnischen Zugehörigkeit, der Intelligenz, der Beziehung der Eltern, der Zufriedenheit oder zum Drogenkonsum erheben lassen.[94] Nicht nur über Facebook, sondern auch auf anderen Social-Media-Plattformen ergeben sich Möglichkeiten, private Details von Nutzern zu erheben. So konnten über eine Farbwertanalyse von auf Instagram hochgeladenen Bildern Rückschlüsse auf die psychische Gesundheit des Nutzers gezogen werden, auch vor einer klinischen Diagnose und mit teilweise besseren Ergebnissen als von Allgemeinärzten.[95] Auch Rückschlüsse auf andere Merkmale einer Person, z. B. die Intelligenz, sind unter Berücksichtigung von Profilbildern möglich.[96] Twitter bietet ebenfalls Möglichkeiten zur Auswertung von Nutzerinformationen. In einer Studie von 2011 konnte mittels einer Untersuchung von Tweets auf die Persönlichkeit der Nutzer geschlossen werden.[97] Diese Wege der Auswertung heben die Mittel des spezifischen Ansprechens von Nutzern auf eine neue Stufe und ermöglichen somit auch eine neue Qualität von Fake News, die auf diese Art in anderen Massenmedien (Printmedien, audiovisuelle Medien, massenhaft verbreitete Speichermedien) bisher nicht möglich war. Um zahlreiche Menschen zu erreichen, ist es ausreichend, zunächst eine geringe Anzahl großer, einflussreicher und vernetzter Accounts anzusprechen – Fake News verbreiten sich wie ein Virus im Netzwerk.[98] Einige der zuvor vorgestellten Wirkmechanismen erreichen im Social-Media-Bereich erst ihre tatsächliche Wirkung bzw. entstammen diesem sogar, z. B. Filterblasen, Smoke-Screening und Misdirecting.

[94] Vgl. Kosinski, M. et al. (2013): Private traits and attributes, S.5803f.
[95] Vgl. Reece, A. und Danforth, C. (2017): Instagram photos reveal predictive markers, S.8f.
[96] Vgl. Wei, X. und Stillwell, D. (2017): How smart does your profile image look, S.39
[97] Vgl. Golbeck, J. et al. (2011): Predicting Personality from Twitter, S.154
[98] Vgl. Andrews, E. (2019): How fake news spreads like a real virus

Es zeigt sich, dass Social Media als Plattform für Nutzer im Internet in Bezug auf Fake News ein besondere Stellung einnehmen, da durch die wachsende Menge an Nutzern und Nutzerinteraktionen und die Verbreitung der Plattformen in der Bevölkerung Fake News schnell, großflächig und zielgerichtet verbreitet werden können. Durch die Möglichkeiten moderner Technologien und die Vernetzung erhalten die Erstellung und Verbreitung von Fake News eine neue Qualität, die mit früheren Mitteln schwer zu erreichen gewesen wäre. Aus diesem Grund wird die weitere Betrachtung von Fake News im Folgenden spezifisch für den Social-Media-Bereich erfolgen.

5.2 Erstellung und Verbreitung von Fake News in Social Media

Fake News können auf verschiedene Arten im Zusammenhang mit Social Media erstellt und verbreitet werden. Zunächst sollen mögliche Ansätze einer automatisierten Erstellung betrachtet werden, um im Anschluss auf die Verbreitungsmöglichkeiten einzugehen.

5.2.1 Automatisierte Erstellung

Häufig werden Fake News als Text verbreitet, teilweise unterstützt durch Bilder oder Videos. Ein solcher Text kann auf verschiedene Arten erstellt werden. Grundsätzlich ist ein manuelles Verfassen von Fake News denkbar, bei dem der Ersteller einen Artikel selbst verfasst. Die Generierung von Fake News, die anschließend von ‚Social Bots' gepostet und verbreitet werden können, muss aber nicht stets manuell erfolgen, sondern kann auch teilweise oder vollständig automatisiert werden. Auf diese Weise kann die Menge der in Umlauf gebrachten Fake News erhöht werden. So können beispielsweise bestehende korrekte Pressemeldungen als Grundlage verwendet und anschließend sprachlich verändert werden.[99] Dadurch wirken Fake News glaubwürdiger, da sie an einen journalistischen Schreibstil angelehnt sind. Ebenso gibt es Technologien, z. B. künstliche Intelligenz, mit denen vollkommen automatisiert nach der Eingabe einiger Stichworte oder vorgegebener

[99] Vgl. Fraunhofer FKIE (2019): Software für die automatisierte Erkennung von Fake News

Textbausteine ein Fake-News-Artikel verfasst werden kann.[100] Mittels eines Algorithmus, der mit großen Mengen an Text aus dem Internet trainiert wurde, können sowohl korrekte Nachrichten als auch Fake News automatisiert generiert werden.[100] Ähnlich funktioniert ‚Grover', ein Tool zur automatisierten Fake-News-Generierung und die Erkennung automatisiert generierter Fake-News-Artikel. Damit können ebenfalls automatisiert anhand vorgegebener Informationen und Wortbausteine (Fake-)News-Artikel erstellt werden.[101] Diese Technologie soll dazu verwendet werden, automatisiert generierte Fake News besser verstehen und erkennen zu können, sie könnte aber auch missbraucht werden.

Nicht nur Text kann benutzt werden, um Fake News zu verbreiten. Sie können auch in Form von Bild, Ton oder Video auftreten. So gibt es beispielsweise Ansätze, die Synthetisierung von Text-zu-Sprache-Audio mittels neuraler Netzwerke zu verbessern.[102] Mithilfe von maschinellem Lernen gelingt es bei sogenannten ‚Deepfakes', realistisch wirkende Videos oder Audiospuren von Personen zu erstellen, die darin beliebige und vom Ersteller ausgewählte Aussagen machen können.[103] Software zur Erstellung von Deepfakes, z. B. das ‚DeepFaceLab', ist im Internet frei verfügbar.[104] Auch die Synthetisierung von Bildern ist bereits in Ansätzen erforscht, so beispielsweise das automatisierte Umwandeln simpler Skizzen in farbige Bilder[105] oder das Generieren von Bildern auf Basis detaillierter Textbeschreibungen.[106] Es zeigt sich, dass sowohl Bild als auch Ton und Videos für Fake News missbraucht werden können. Dies ist aber aufwändiger als das alleinige Erstellen von Text. Daher ist davon auszugehen, dass Textformate den größten Anteil der produzierten Fake News ausmachen.

[100] Vgl. Knight, W. (2019): An AI that writes convincing prose risks mass-producing fake news
[101] Vgl. Zellers, R. et al. (2019): Defending Against Neural Fake News, S.3ff.
[102] Vgl. Arik, S. et al. (2017): Deep Voice, S.1f.; vgl. Kalchbrenner, L. et al. (2018): Efficient Neural Audio Synthesis, S.1
[103] Vgl. Laaff, M. (2019): Hello, Adele – bist du's wirklich
[104] Vgl. Perov, I. et al. (2020): DeepFaceLab, S.1f.
[105] Vgl. Isola, P. et al. (2016): Image-to-Image Translation with Conditional Adversarial Networks, S.1f.
[106] Vgl. Zhang, H. et al. (2017): StackGAN, S.5908f.

5.2.2 Automatisierte Verbreitung

Sind geeignete Fake News erstellt worden, müssen sie verbreitet werden, um mögliche Empfänger zu erreichen und wirksam zu sein. Die Verbreitung auf Social-Media-Plattformen wie Facebook oder Twitter kann auf verschiedene Arten erfolgen. Ein Post auf einer solchen Plattform kann entweder die Fake News selbst enthalten oder darauf verlinken. Sind die Falschnachrichten auf einer separaten Webseite zu finden, orientiert sich das Design der Fake-News-Webseite häufig an jenem seriöser Nachrichtenseiten.[107] Sollte eine Nachricht zu lang sein, um auf Social Media angemessen dargestellt werden zu können, kann mit einer Überschrift Neugier erzeugt und mit einem Link auf eine Webseite verwiesen werden. Mittels verkürzter URLs ist es möglich, die Identität einer Webseite zu verschleiern.[108] Dadurch kann sie beispielsweise als seriöses Nachrichtenportal auftreten. Da der Fokus hier vor allem auf dem Social-Media-Bereich liegt, werden hauptsächlich Social-Media-interne Nachrichten betrachtet. Abgesehen von einer rein manuellen Verbreitung, bei der jede Nachricht von einem Menschen verfasst und verbreitet wird, können auch Algorithmen eingesetzt werden. Diese können genutzt werden, um automatisiert Nachrichten auf Social-Media-Plattformen zu erstellen, diese zu verbreiten oder Interaktion zu simulieren. Während eine manuelle Verbreitung von Fake News prinzipiell denkbar ist, erscheint die Nutzung von zusätzlicher Technologie zur schnelleren, effizienteren und intensiveren Verbreitung sinnvoll. Dafür werden in der Regel Social Bots verwendet.[109] Das Wort ‚Social' soll dabei die Imitation menschlichen Verhaltens ausdrücken und stammt nicht von ‚Social Media' ab[110], auch wenn Bots dieser Art dort verbreitet sind. ‚Bot' ist die Kurzform von ‚Robot'.[111] Der Begriff ‚Social Bot' kann grundsätzlich sowohl gutartige als auch bösartige Bots beschreiben.[112] Diese Einordnung hängt von der jeweiligen Zielsetzung des Bots ab. Gutartige Bots können beispielsweise informieren und legitime Nachrichten posten, während bösartige Bots dazu genutzt werden können, Fake News zu verbreiten. Stieglitz et al. schlagen zusätzlich das Einführen einer ‚neutralen' Kategorie beispielsweise für humoristische Bots vor.[113] In Abbildung 2 sind

[107] Vgl. Ruddick, G. (2017): Experts sound alarm over news websites' fake news twins
[108] Vgl. Chhabra, S. et al. (2011): Phi.sh/$oCiaL, S.92
[109] Vgl. Hegelich, S. (2016): Invasion der Meinungs-Roboter, S.2
[110] Vgl. Akademische Gesellschaft (2018): How powerful are Social Bots, S.1
[111] Vgl. Howard, P. et al. (2018): Algorithms, bots, and political communication, S.82
[112] Vgl. Brachten, F. et al. (2018): Threat or Opportunity, S.3
[113] Vgl. Stieglitz, S. et al. (2017): Do Social Bots Dream of Electric Sheep, S.4f

die grundlegenden Kategorien von Social Bots mit zwei beispielhaften Aufgaben je Kategorie aufgeführt.

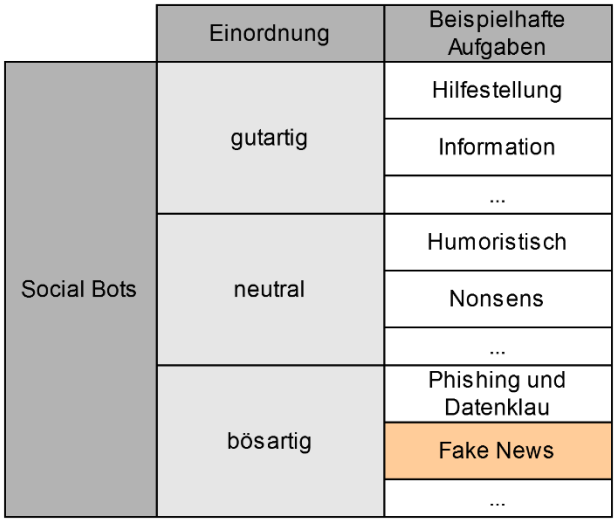

Abbildung 2: Einordnung von Social Bots

Laut einer Untersuchung waren 2019 für 62,8 % aller Webseitenaktivitäten Menschen, für 13,1 % gutartige Bots und für 24,1 % aller Webseitenaktivitäten sogenannte ‚Bad Bots', also bösartige Bots, verantwortlich.[114] Bösartige Social Bots können von Assistenz-Bots, Trollen, Cyberangriffen und Spam-E-Mails bezüglich ihrer Zielsetzung abgegrenzt werden.[115] Während bösartige Social Bots durch den Versuch der Einflussnahme (E), einen implementierten Algorithmus (A) und das Vortäuschen menschlicher Identität (I) charakterisiert werden können, trifft dies auf die anderen Internetphänomene nicht zu. Assistenz-Bots können beispielsweise lediglich mit Aspekt A, Trolle mit E, Cyberangriffe mit der Kombination EA und Spam-E-Mails mit AI charakterisiert werden. Die technischen Grundlagen dieser Phänomene sind ähnlich.

Generell sind Social Bots Computerprogramme, die abhängig von ihrer Funktion, ihrer Fähigkeit und ihrem Design in ihrer Größe variieren und auf Social-Media-Plattformen eingesetzt werden können, um verschiedene Aufgaben auszuführen.[116] Dabei simulieren sie menschliches Verhalten und können dafür künstliche Intelligenz, Big-Data-Analysen, Datenbanken und andere Programme verwenden.[116] Die Qualität von Social Bots kann

[114] Vgl. Roberts, E. (2020): Bad Bot Report 2020
[115] Vgl. Kind, S. et al. (2017): Social Bots, S.4
[116] Vgl. Office of Cyber and Infrastructure Analysis (2018): Social Media Bots Overview

unterschiedlich sein. Während einfache Bots beispielsweise lediglich Schlüsselwörter erkennen oder Inhalte aus dem Internet posten oder retweeten, gelingt es komplexeren Bots, Nachrichten zu analysieren und Unterhaltungen zu führen.[115] Ein Großteil der Social Bots ist eher simpel[115] und existiert zu unterstützenden Zwecken oder bietet eingeschränkt automatisierte Module.[117] Social Bots sind bisher nicht weithin verfügbar und bieten zumeist lediglich einfache Funktionen wie Liken, Teilen oder Folgen.[118] In einer Untersuchung von Assenmacher et al. war der Großteil des verfügbaren Codes für Social Bots für Telegram, Twitter, Facebook und Reddit konzipiert.[118] Vor allem komplexere Social Bots können Konversationen führen und wie menschliche Nutzer wirken.

Bots können oftmals gegen Bezahlung in großer Stückzahl erworben oder auch selbst programmiert werden. Die dazugehörigen Accounts können ebenfalls online gekauft werden.[119] Die Urheber dieser Verbreitung sind häufig schwierig oder nicht nachverfolgbar.[115]

Ein vollständig automatisiertes Betreiben eines Social Bots ohne menschliche Interaktion auf Social Media wird als ‚Automated Social Media Bot' und eine zusätzliche Unterstützung mit menschlicher Interaktion als ‚Semi-automated Social Media Bot' bezeichnet.[116] Alternativ findet sich für eine vollständige Automatisierung der Begriff ‚Bot' und für eine teilweise Automatisierung mit menschlicher Interaktion die Bezeichnung ‚Cyborg'.[120] Der Zugriff von Social Bots auf Social-Media-Plattformen erfolgt von einem Server aus[121] über Application-Programming-Interfaces (API).[122] Umso einfacher der Zugriff auf ein API gestaltet ist, desto mehr Bots sind auf einer Plattform zu erwarten. Das führt dazu, dass auf Social-Media-Plattformen wie Twitter oder Instagram mehr Bots eingesetzt werden als beispielsweise auf Facebook.[119]

Social Bots haben auf Social Media verschiedene klassische Verhaltensmuster bzw. Aufgaben. So werden sie unter anderem für das Erhöhen der Popularität von Inhalten genutzt, z. B. indem Posts oder Accounts geliked, diesen gefolgt oder sie weiterverbreitet werden.[116] Auf die gleiche Weise können Hashtags bzw. bestimmte Themen populär gemacht oder auch bereits populäre Hashtags für die eigene Reichweite genutzt werden. Andere

[117] Vgl. Assenmacher, D. et al. (2020): Demystifying Social Bots, S.1
[118] Vgl. Assenmacher, D. et al. (2020): Demystifying Social Bots, S.5f
[119] Vgl. Hegelich, S. (2016): Invasion der Meinungs-Roboter, S.2f.
[120] Vgl. Chu, Z. et al. (2012): Detecting Automation of Twitter Accounts, S.811
[121] Vgl. Howard, P. et al. (2018): Algorithms, bots, and political communication, S.85
[122] Vgl. Kind, S. et al. (2017): Social Bots, S.6

Nutzer werden kontaktiert oder es wird, um die Reichweite zu vergrößern, ein Netzwerk (mit möglichst populären menschlichen Nutzern darin) aufgebaut.[123] Auch das Kopieren von Identitäten echter Personen ist möglich, z. B. mittels eines ähnlichen Namens, kopierter Bilder und durch das Kopieren des Verhaltens der Person durch Interaktion mit deren Freunden und die Imitation des Postingverhaltens.[123] Durch ihr Auftreten, z. B. mit Profilbild (automatisiert bezogen aus dem Internet)[123], Beschreibung, realistischen Postings und ohne besondere Kennzeichnung, können Social Bots häufig schwer von menschlichen Nutzern unterschieden werden.[115] Da sie durch ihre Täuschung das öffentliche Stimmungsbild verzerren können, werden sie auch als ‚Influence-Bots' bezeichnet.[124] Selbst ein ungeordneter massenhafter Einsatz von Bots (z. B. durch Teilen auf Facebook oder Hashtags auf Twitter) kann die Wahrnehmung von Plattformalgorithmen für aktuell beliebte Inhalte manipuliert werden, die dann zahlreichen Nutzern angezeigt werden.[125]

Um funktionieren zu können, sind Social Bots auf eine gewisse technische Infrastruktur angewiesen. Diese besteht in der Regel aus einer Kombination aus einem Profil auf einer Social-Media-Plattform, technischen Möglichkeiten zur Automatisierung des Accountverhaltens durch das API oder anderen Interaktionsmöglichkeiten und Algorithmen als Grundlage.[126] Um mit Menschen kommunizieren zu können, werden bei Social Bots Technologien eingesetzt, die auch bei klassischen Chatbots zu finden sind. Der Begriff ‚Chatbot' setzt sich aus ‚Chat' und ‚Robot' zusammen und bezeichnet ein automatisiertes Onlinekommunikationssystem.[127] Durch diese Kommunikationsmöglichkeit verbessern sich die Täuschungsmöglichkeiten eines Social Bots. Naheliegend ist dabei die Konversation mit Menschen, allerdings können auch einfachere Aufgaben, z. B. das Weiterverbreiten existierender Posts, durchgeführt werden. Chatbots bestehen aus drei Komponenten: Responder, Classifier und Graphmaster.[128] Während der Responder das Interface zwischen Nutzer und Bot darstellt, wandelt der Classifier den erhaltenen Input so um, dass der Bot ihn verarbeiten kann. Der Graphmaster ist für die Speicherung und Sortierung von Informationen zuständig.[128] Die Generierung von Output auf Basis des Inputs

[123] Vgl. Ferrara, E. et al. (2016): The Rise of Social Bots, S.99f.
[124] Vgl. Subrahmanian, V. et al. (2016): The DARPA Twitter Bot Challenge, S.38
[125] Vgl. Hegelich, S. (2016): Invasion der Meinungs-Roboter, S.4
[126] Vgl. Assenmacher, D. et al. (2020): Demystifying Social Bots, S.2
[127] Vgl. Luber, S. und Litzel, N. (2018): Was ist ein Chatbot
[128] Vgl. Abdul-Kader, S. und Woods, J. (2015): Survey on Chatbot Design Techniques, S.73

findet im Responder statt, entweder durch eigenständige Generierung oder mittels einer Abfrage aus einer Datenbank.[129] Generiert er Output, wird ein Chatbot für gewöhnlich mit Texten, z. B. von einer bestimmten Person oder aus einem Buch, trainiert.[130] Eine Generalisierung gelingt damit aber häufig nicht, sondern der Stil des Originalautors wird imitiert.[130] Assenmacher et al. Schätzen die Entwicklung von Social Bots mit einfachen (Folgen, Liken, Teilen) und anspruchsvolleren Aufgaben (das Simulieren menschlichen Verhaltens ohne Interaktion mit anderen Menschen) als realisierbar ein. Ein Social Bot, der glaubwürdige Unterhaltungen mit Menschen führt, scheint aber noch nicht existent zu sein.[130] Der technologische Fortschritt sei noch nicht weit genug vorangeschritten, um wirklich intelligente Bots zu realisieren.

Mehrere Bots können in Botnetzen Social-Media-Plattformen infiltrieren, um beispielsweise die Reichweite von Inhalten zu erhöhen oder Nutzerdaten zu sammeln.[131] Damit soll ein menschliches Freundenetzwerk simuliert werden.[132] Der Begriff ‚Botnetz' ist eine Kombination aus ‚Robot' und ‚Netzwerk' und beschreibt Algorithmen, die über mehrere Geräte hinweg kommunizieren, um eine Aufgabe zu erledigen.[133] Diese Netze können lange inaktiv sein, bis sie eingesetzt werden. Sie werden häufig von einem zentralen Verwalter (‚Botherder') mittels eines Steuerungstools (‚Botmaster') gesteuert.[134] Um Anti-Bot-Algorithmen von Plattformen wie z. B. Facebook nicht aufzufallen, integrieren sich die Bots zu Beginn ähnlich wie ein neuer Nutzer, der sich nach und nach vernetzt.[134] In einer Studie mit 120 automatisierten Social Bots auf Twitter hatte die Plattform nach einem Monat lediglich 31 % identifiziert und entfernt.[135] Dies betraf vor allem die zuletzt angelegten Bots. Simulieren Bots zusätzlich noch menschliches Verhalten, indem sie beispielsweise Pausen machen oder ihre Nachrichten jeweils anpassen[136], kann eine Entdeckung noch erschwert werden. Die Nachrichten eines Bots erscheinen aufgrund seiner falschen menschlichen Identität als glaubwürdiger. Sicherheitsmaßnahmen wie Captchas können umgangen werden, indem verschiedene Maßnahmen ergriffen werden. Mit Tech-

[129] Vgl. Assenmacher, D. et al. (2020): Demystifying Social Bots, S.4
[130] Vgl. Assenmacher, D. et al. (2020): Demystifying Social Bots, S.10f.
[131] Vgl. Office of Cyber and Infrastructure Analysis (2018): Social Media Bots Overview
[132] Vgl. Hegelich, S. und Janetzko, D. (2016): Are Social Bots on Twitter Political Actors, S.582
[133] Vgl. Howard, P. et al. (2018): Algorithms, bots, and political communication, S.83
[134] Vgl. Boshmaf, Y. (2013): Design and Analysis of a Social Botnet, S.557ff.
[135] Vgl. Freitas, C. et al. (2016): Socialbot infiltration strategies in the Twitter social network, S.7
[136] Vgl. Stieglitz, S. et al. (2017): Do Social Bots Dream of Electric Sheep, S.4f

nologien wie Machine Learning oder Buchstabenerkennung, dem Einsatz von menschlichen Nutzern (etwa durch Bezahlung oder Täuschen anderer Plattformnutzer) oder der direkten Bekämpfung der Sicherheitsmaßnahmen (z. B. durch Wiederbenutzung bereits existierender Session-IDs oder durch das Lösen von Hashes) ist dies möglich.[137] Ein solches Botnetz kann auch dazu genutzt werden, Gemeinschaften im Internet (Communities) zu infiltrieren und im Laufe der Zeit Vertrauen aufzubauen, das im Anschluss genutzt werden kann, um beispielsweise an Informationen zu gelangen, Meinungen zu beeinflussen oder die Nutzer zu Handlungen aufzurufen.[137] Das Folgen von Nutzern einer gemeinsamen Interessensgruppe ist ein vielversprechenderer Ansatz als die zufällige Auswahl von Nutzern.[138] Ist eine Community infiltriert, können zahlreiche menschliche Nutzer angegriffen werden. Umso mehr gemeinsame Freunde ein Social Bot mit einem Nutzer hat, desto höher ist die Wahrscheinlichkeit, dass der menschliche Nutzer eine Freundschaftsanfrage des Social Bots annimmt.[139] Ebenso ist die Wahrscheinlichkeit höher, dass ein menschlicher Nutzer die Freundschaftsanfrage eines fremden Social Bots akzeptiert, wenn der Nutzer bereits zahlreiche Freunde auf der Plattform hat.[139] Social Bots mit einem hohen Aktivitätslevel erreichen dabei häufig eine größere Popularität als weniger aktive Bots.[140] Bei einer Untersuchung von Stieglitz et al. generierten Bots neue Follower mit jedem zweiten Tweet.[141] Darin kann ein möglicher Grund für hohe Botaktivitäten liegen. Da Nutzer Nachrichten von Accounts in ihrem Netzwerk bezüglich der Glaubwürdigkeit weniger anzweifeln, besonders, wenn sie dem Account folgen oder dieser sich anderweitig bereits in ihrem Netzwerk befindet, können Fake News einfacher verbreitet werden.[142]

Durch die Nutzung von Machine Learning ist es möglich, automatisiert für eine große Menge an Nutzern den ‚optimalen' Social Bot zu erstellen, der die Wahrscheinlichkeit der Interaktion erhöht. Seymour und Tully gelang es 2016 auf Twitter, eine automatisierte Spear-Phishing-Attacke, also den gezielten Angriff auf einen einzelnen Nutzer (durch Analyse von Nachrichten, Retweets und Follows), in großem Umfang durchzuführen.[143] Durch unterschiedliche Inhalte und Postingzeiten, abgestimmt auf den jeweiligen Nutzer,

[137] Vgl. Boshmaf, Y. (2013): Design and Analysis of a Social Botnet, S.558f.
[138] Vgl. Freitas, C. et al. (2016): Socialbot infiltration strategies in the Twitter social network, S.14
[139] Vgl. Boshmaf, Y. (2013): Design and Analysis of a Social Botnet, S.566
[140] Vgl. Freitas, C. et al. (2016): Socialbot infiltration strategies in the Twitter social network, S.12
[141] Vgl. Stieglitz, S. et al. (2017): Do Social Bots (Still) Act Different to Humans, S.391f.
[142] Vgl. Morris, M. et al. (2012): Tweeting is Believing, S.444
[143] Vgl. Seymour, J. und Tully, P. (2016): Weaponizing Data Science for Social Engineering, S.4f.

konnte so die Wahrscheinlichkeit einer Entdeckung durch Twitter vermindert werden. Der Erfolg des Spear-Phishings wurde dabei daran gemessen, ob der angesprochene Nutzer auf einen beigefügten Link klickte. Im Gegensatz zu klassischen Phishing-Attacken mit Erfolgsquoten von 5–14 % klickten bei diesem automatisierten Ansatz zwischen 33 % und 60 % der Nutzer den Link an.[144] Abgesehen von Spear-Phishing ist auch die automatisierte Generierung von für einen Nutzer besonders glaubwürdigen Fake News denkbar. Allein durch das Liken und Teilen von Informationen können Social Bots Fake News in großem Ausmaß verbreiten.[145]

Es zeigt sich, dass Social Bots, vor allem auch vernetzt in Botnetzen, Erstellern die Möglichkeit bieten, Fake News insbesondere auf Social Media effektiv zu verbreiten. Um dieser Verbreitung entgegenzuwirken, existieren verschiedene Erkennungsansätze im Social-Media-Bereich. Da bei der Erstellung von Fake News Automatisierung genutzt wird, sollte diese für die Erkennung ebenfalls bedacht werden.

[144] Vgl. Seymour, J. und Tully, P. (2016): Weaponizing Data Science for Social Engineering, S.7
[145] Vgl. Lazer, D. et al. (2018): The science of fake news, S.1095

6 Erkennung von Fake News in Social Media

Bereits Menschen fällt es teilweise schwer, Fake News korrekt zu identifizieren. Eine automatisierte Unterstützung kann hilfreich sein, um große Mengen an Nachrichten, die eventuell automatisiert in Umlauf gebracht werden, zeitnah zu sichten und einzuordnen.[146] Problematisch ist, dass Informationen nicht nur vollständig richtig oder vollständig falsch sein können, sondern auch in einem Bereich dazwischen liegen können. Der Übergang zwischen Fake News und ähnlichen Formaten wie Satire[147] kann ebenfalls fließend sein. Ebenso kann es sich um eine Meinung handeln, die nicht objektiv überprüfbar ist.[148] Die Automatisierung des Prüfungsvorgangs ist somit eine Herausforderung. Nicht nur eine vollständige Automatisierung ist denkbar, sondern auch die Unterstützung menschlicher Prüfer durch diese.[146] Die Kernelemente einer automatisierten Faktenprüfung sind nach Graves zunächst die Identifikation überprüfbarer Inhalte, anschließend deren Verifizierung mittels Faktenüberprüfungen und schließlich die darauffolgende Korrektur.[146]

Um diesen Prozess und seine Ergebnisse zugänglicher zu machen, existieren bereits vorgefertigte Tools bzw. Ideen. Um Privatpersonen die Nutzung eines Tools zur Faktenprüfung zu ermöglichen, schlagen Aldwairi und Alwahedi vor, eine herunterladbare Software bereitzustellen, die Suchergebnisse prüft und filtert, bevor diese dem Nutzer angezeigt werden.[149] Die Anzeige der Ergebnisse von Faktenüberprüfungen mittels einer App ist auch denkbar[146], ebenso der Einsatz von Browsererweiterungen. Für die Plattform Twitter existiert eine Browsererweiterung namens ‚TweetCred', die anhand der durch einen Post verfügbaren Daten die Glaubwürdigkeit einer Nachricht einschätzt.[150] Damit andere Nutzer davon erfahren, wenn Fake News festgestellt wurden, ist neben einer klassischen Gegendarstellung als zusätzliche Nachricht oder Kommentar ebenso die Einführung weiterer Darstellungsmaßnahmen denkbar: beispielsweise Widgets, die den Nutzer informieren, oder direkte Einblendungen und Verweise auf Faktenchecks unter den Suchergebnissen von Google.[151]

---

[146] Vgl. Graves, L. (2018): Understanding the Promise and Limits, S.2ff.
[147] Vgl. Horne, B. und Adali, S. (2017): This Just In, S.6
[148] Vgl. Hassan, N. et al. (2015): The Quest to Automate Fact-Checking, S.3
[149] Vgl. Aldwairi, M. und Alwahedi, A. (2018): Detecting Fake News in Social Media Networks, S.216; vgl. Graves, L. (2018): Understanding the Promise and Limits, S.5
[150] Vgl. Gupta, A. et al. (2014): TweetCred, S.228f.
[151] Vgl. Adair, B. et al. (2017): Progress Toward „the Holy Grail", S.3f.

Da Ergebnisse dieser Art erst präsentiert werden können, wenn sie vorliegen, müssen Nachrichten auf ihre Korrektheit geprüft werden. Häufig haben sich unabhängige Personen oder Personengruppen bereits zum Ziel gesetzt, die aktuelle Berichterstattung zu überprüfen und Fake News aufzudecken, auch automatisiert. Jedoch kann nicht nur durch externe Personen eine Überprüfung erfolgen. Auch Social-Media-Plattformen selbst (z. B. Facebook oder Twitter) geben an, gegen Fake News vorzugehen, stellen aber keine konkreten Maßnahmen vor, wie die Erkennung vorgenommen werden soll.[152] Es werden dabei verschiedene Systeme eingesetzt, die unterschiedliche Erkennungsansätze nutzen können. In der Regel basieren Systeme zur automatisierten Fake-News-Erkennung auf Machine Learning[153], Deep Learning[154] oder regelbasierten Methoden[155]. Da Deep Learning für gewöhnlich eine größere Anzahl an wahren Nachrichten und Fake News zu einem Thema als Datengrundlage voraussetzt, auf die zumeist schwer zurückgegriffen werden kann, werden häufig auch Machine Learning-Methoden verwendet, wobei dabei weniger Nachrichten benötigt werden, im Gegenzug aber die zu analysierenden Merkmale vorgegeben werden.[156]

Eine Herausforderung ist es oftmals, eine ausreichend große Datenbasis für das Training eines Algorithmus zu erhalten, da sowohl zahlreiche wahre Nachrichten als auch Fake News benötigt werden.[156] Eine Datenbasis, auf der ein Algorithmus aufbauen kann, kann z. B. eine Kombination aus klassifizierten Fake News und wahren Nachrichten sein, also eine Bibliothek bereits durchgeführter Faktenchecks.[146] Damit andere Forscher auf diese zurückgreifen können, werden klassifizierte Texte als Datenbasis von verschiedenen Quellen bereitgestellt.[157] Grundlegend werden dabei für die Datenbasis entweder tatsächlich existierende Artikel verwendet und mit einem Label versehen[157], korrekte Nachrichten eigenständig so umgewandelt, dass dabei Fake News entstehen[158], oder vollständig fiktive Texte erstellt.[159] Eine unzureichende Datenbasis kann zu dürftigen Ergebnissen

[152] Vgl. Weedon, J. et al. (2017): Information Operations and Facebook, S.5; vgl. Roth, Y. und Pickles, N. (2020): Updating our approach to misleading information
[153] Vgl. Reis, J. et al. (2019): Supervised Learning for Fake News Detection, S.78f.; vgl. Yang, S. et al. (2019): Unsupervised Fake News Detection on Social Media, S.5644ff.
[154] Vgl. Popat, K. et al. (2018): DeClarE, S.1ff.; vgl. Yang, Y. et al. (2018): TI-CNN, S.1ff.; vgl. Ruchansky, N. et al. (2017): CSI, S.797ff.
[155] Vgl. Feng, S. (2012): Syntactic stylometry for deception detection, S.171ff.
[156] Vgl. Hen, J. (2020): Automatisierte Wahrheitssuche
[157] Vgl. Wang,W. (2017): „Liar, Liar Pants on Fire“, S.1; vgl. Reis, J. et al. (2019): Supervised Learning for Fake News Detection, S.77
[158] Vgl. Pérez-Rosas, V. et al. (2018): Automatic Detection of Fake News, S.3393
[159] Vgl. Zellers, R. et al. (2019): Defending Against Neural Fake News, S.3ff.

bezüglich der Genauigkeit der Klassifikation führen.[160] Dabei sollte darauf geachtet werden, dass auch bekannte Fake-News-Quellen wahre Nachrichten oder Quellen wahrer Nachrichten Fake News veröffentlichen können. Dies sollte die Datenbasis nicht verunreinigen.[160]

Um Fake News automatisiert zu bekämpfen, existieren unterschiedliche Ansätze. Da eine Nachricht in der Regel über Inhalt und eine Quelle verfügt und andere Personen auf diese reagieren, können aus drei verschiedenen Bereichen Rückschlüsse darauf gezogen werden, ob es sich bei einer Nachricht um Fake News handelt oder nicht. Der erste Ansatz ist es, den Text der Nachricht zu analysieren. Dabei besteht die Möglichkeit, Aussagen im Text auf ihre Korrektheit zu überprüfen oder den benutzten Schreibstil zu untersuchen. Ein zweiter Ansatz besteht darin, die Quelle zu überprüfen, die die Nachricht veröffentlicht hat. Die Quelle muss dabei nicht zwingend die Nachricht erstellt haben. Der dritte Ansatz ist es, die Reaktion anderer Nutzer auf eine Nachricht zu analysieren, um Rückschlüsse auf die Korrektheit des Inhalts ziehen zu können. Es wird also die Umgebung der Nachricht und teilweise auch jene der Quelle betrachtet. Wie in Abbildung 3 dargestellt ist, wird bei Betrachtung des Inhalts einer Nachricht der eigentliche Kern dieser analysiert, während die Überprüfung der Quelle oder der Umgebung sich von der eigentlichen Nachricht entfernt und daher indirekter ist.

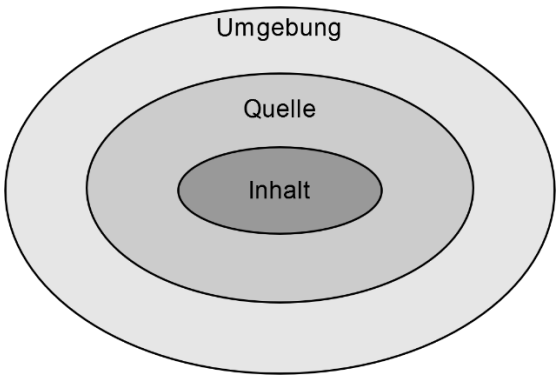

Abbildung 3: Betrachtung von Inhalt, Quelle oder Umgebung

Der Ansatz der Quellenbetrachtung umschließt die einzelne Nachricht, aber auch andere Faktoren in Bezug auf die veröffentlichende Quelle. Noch weiter greift die Herangehensweise der Umgebungsbetrachtung. Dabei wird das Umfeld von Nachricht und Quelle betrachtet. Es können dabei auch andere Nutzer, nicht nur die Quelle, einbezogen werden.

[160] Vgl. Kapusta, J. und Obonya, J. (2020): Improvement of Misleading and Fake News Classification, S.8

Eine Kombination von Ansätzen ist ebenso möglich. Im Folgenden sollen die drei Ansätze detaillierter beschrieben und im Anschluss daran auf Kombinationen davon eingegangen werden.

6.1 Betrachtung des Inhalts

Hinsichtlich der inhaltlichen Merkmale einer Nachricht werden hauptsächlich die Überschrift und der Textkörper betrachtet, aber auch weitere Elemente, etwa Bilder oder Videos, können einbezogen werden.[161]

Während bisher beispielsweise eine direkte automatisierte Überprüfung des Inhalts von Bildern lediglich schwierig möglich war, können Rückschlüsse auf den Wahrheitsgehalt durch Informationen aus dem Umfeld oder dem Begleittext gezogen werden.[162] Durch die Einbeziehung weiterer statistischer und visueller Merkmale kann auch in diesem Bereich das Bild an sich berücksichtigt werden. So gehen z. B. Jin et al. davon aus, dass eine Menge von Bildern zu einem realen Geschehnis heterogener ist als eine Menge von Bildern zu Fake News.[163] Ebenso konnten in einer Untersuchung von Yang et al. über die vorliegende Auflösung eines Bildes und die Anzahl der darauf sichtbaren Gesichter Rückschlüsse gezogen werden.[164] Bilder, die mit Fake News in Verbindung gebracht werden konnten, enthielten dabei durchschnittlich weniger Gesichter und hatten eine geringere Auflösung. Da das Erstellen gefälschter Bilder allerdings einen größeren Aufwand darstellt und automatisiert schwieriger umsetzbar ist als das Erstellen von textuellen Fake News, ist davon auszugehen, dass die größte Menge an Fake News in Social Media durch Text mitgeteilt wird.

Sind externe Links an die Nachricht angehängt, können diese ebenfalls einen Hinweis auf die automatisierte Verbreitung von Nachrichten und damit auf Fake News geben. So ist die Ratio von externen Links in Nachrichten und der Gesamtzahl der Nachrichten bei Bots höher als bei Menschen.[165]

[161] Vgl. Shu, K. et al. (2017): Fake News Detection on Social Media, S.5f.
[162] Vgl. Gupta, A. et al. (2013): Faking Sandy, S.733
[163] Vgl. Jin, Z. et al. (2017): Novel Visual and Statistical Image Features, S.599ff.
[164] Vgl. Yang, Y. et al. (2018): TI-CNN, S.4
[165] Vgl. Chu, Z. et al. (2012): Detecting Automation of Twitter Accounts, S.820

Grundlegend beschäftigt sich der Ansatz, den Text der Nachricht auf Fehler und Fälschungen zu überprüfen, mit den Aussagen an sich oder mit dem Schreibstil.

6.1.1 Überprüfung des Inhalts von Aussagen

Die Überprüfung der Aussagen einer Nachricht ist naheliegend, da dadurch der Wahrheitsgehalt direkt kontrolliert werden kann.

Neben der Vorlage bei menschlichen Prüfern gibt es Ansätze der automatisierten Kontrolle. Um den Wahrheitsgehalt eines Textes zu prüfen, müssen zunächst prüfbare Aussagen identifiziert werden. Dafür existiert beispielsweise das Tool ‚ClaimBuster' von Adair, B. et al.[166] Damit wird ein bereitgestellter Text durchsucht und die Sätze darin nach deren Überprüfungsnotwendigkeit gewichtet. Damit dies möglich ist, wurde ein Algorithmus mit einer großen Anzahl von von Menschen klassifizierten Sätzen trainiert. Subjektiv oder meinungsbezogen formulierte Sätze werden niedrig gewichtet und weisen daher eine geringere Notwendigkeit zur Überprüfung auf. Die Behauptung von Fakten erhält hingegen ein eher hohes Gewicht. [167] Eine Herausforderung bei der automatisierten Prüfung von Aussagen ist es, zu unterscheiden, ob die Überprüfung einer Aussage für die Öffentlichkeit relevant ist.[168] Einen Hinweis, ob eine Nachricht besonders überprüfungswürdig ist oder nicht, kann eine Überprüfung der innerhalb dieser Nachricht dargestellten Haltung (Stance-Detection) liefern, indem abgeglichen wird, ob innerhalb einer Nachricht Widersprüche vorhanden sind, z. B. zwischen der Überschrift und dem tatsächlichen Inhalt des Textkörpers.[169]

Sind überprüfungswürdige Aussagen identifiziert worden, kann die Korrektheit des Inhalts kontrolliert werden. Dabei besteht die Möglichkeit, expertenorientiert, Crowdsourcing-orientiert oder automatisiert vorzugehen.[161] Für eine expertenortientierte Bewertung werden menschliche Experten herangezogen, die manuell den Wahrheitsgehalt einer Nachricht bewerten. Diese Methode ist aufgrund der Bindung an bestimmte Experten langsam und zeitaufwändig. In Projekten wie PolitiFact[170]oder Snopes[171] wird auf diese

[166] Vgl. Adair, B. et al. (2017): Progress Toward „the Holy Grail", S.1f.
[167] Vgl. Hassan, N. et al. (2017): ClaimBuster, S.1945f.
[168] Vgl. Hassan, N. et al. (2015): The Quest to Automate Fact-Checking, S.3
[169] Vgl. Hanselowski, A. et al. (2018): A Retrospective Analysis of the Fake News Challenge, S.1f.
[170] Vgl. Holan, A. (2018): The Principles of the Truth-O-Meter
[171] Vgl. Snopes (2020): Snopes.com follows all industry guidelines for transparency in reporting

Weise vorgegangen. Die Einbeziehung von zahlreichen anderen Nutzern (Crowdsourcing), die keinen Expertenstatus haben, ist ebenfalls eine Möglichkeit, mögliche Fake News zu überprüfen. Aus allen getätigten Bewertungen wird der durchschnittlich von den Nutzern angegebene Wahrheitsgehalt berechnet und für andere Nutzer angezeigt. Ein Beispiel für ein solches Vorgehen ist Fiskkit.[172] Dabei kann die Vorlage bei einem Menschen auch mit einer vorhergehenden automatisierten Datenaufbereitung einhergehen. Es kann notwendig sein, die Aussagen so umzuformen, dass sie einfacher zu prüfen sind.[166] Damit sind sie im Anschluss geeigneter, von weiteren Systemen verarbeitet zu werden. ClaimBuster benutzt den Ansatz, Fragen umzuformulieren und an Systeme, die dafür geeignet sind, Fragen zu beantworten (z. B. Wolfram Alpha und Google), zu schicken. Umgebende Sätze in Texten dieser Systeme werden als Kontext miteinbezogen und anschließend dem Nutzer vorgelegt.[167] Um eine vollständige Automatisierung zu erreichen, müsste das Ergebnis anschließend automatisiert weiterverarbeitet werden. Zu überprüfende Aussagen können automatisiert mittels einer Suche im Internet[173] oder einer Datenbank, die vergangene Faktenchecks[174] enthält, überprüft werden. Sind Links in einer Nachricht enthalten, können die dahinterliegenden Webseiten mit Blacklists geprüft werden.[165]

Die Nutzung eines ‚Knowledge-Graphs' kann die automatisierte, inhaltliche Websuche vereinfachen. Wie sehr eine Aussage unterstützt werden kann, wird approximiert, indem der Verbindungsweg der Knoten einer Aussage auf einem Graphen nachvollzogen wird. Die Subjekte und Objekte eines Satzes sind die Knoten, während die Prädikate die Verbindungswege bilden.[175] Dadurch können alle Subjekte und Objekte zahlreicher Aussagen in einem Knowledge-Graph-Netzwerk miteinander verbunden werden. Ein Netzwerk wird beispielsweise durch das Auslesen zahlreicher Infoboxen aus Wikipedia, also von öffentlich zugänglichen Quellen, aufgebaut. Gibt es eine neue überprüfbare Aussage, wird der zugehörige Pfad im Netzwerk gesucht. Ist eine solche Verbindung bereits vorhanden oder ist sie durch kurze Wege zwischen den Knoten erreichbar, wird eine Aussage als wahr eingeschätzt. Weniger spezifische Subjekte bzw. Objekte (z. B. ‚Deutschland')

[172] Vgl. Fiskkit (2020): What is "Fisking"
[173] Vgl. Popat, K. et al. (2018): DeClarE, S.1ff.
[174] Vgl. Full Fact (2020): Automated Fact Checking
[175] Vgl. Ciampaglia, G. et al. (2015): Computational Fact Checking from Knowledge Networks, S.2f.

unterstützen eine Aussage dabei weniger als spezifischere (z. B. ‚Friedrich-Schiller-Universität Jena').[175] Eine weitere mögliche Nutzung von Knowledge-Graphen zur automatisierten Überprüfung von Inhalten schlagen Shi und Weninger vor.[176] Aussagen werden Entitäten zugeordnet (z. B. ‚Jena' der Entität ‚Stadt'). Auch hier werden Sätze nach Subjekt, Prädikat und Objekt aufgeteilt. Indem ein anderer Pfad mittels alternativer Entitätspaare im Netzwerk gesucht wird, wird versucht, festzustellen, ob eine Aussage der Wahrheit entspricht.[176]

Eine Schwierigkeit bleibt die Überprüfung von Aussagen, die nicht eindeutig falsch oder korrekt sind, sondern beispielsweise Halbwahrheiten enthalten[177] oder über keine simple Satzstruktur verfügen.[176] Daher kommt die rein automatisierte Überprüfung vor allem bei zahlenbasierten Aussagen zum Einsatz, die Überprüfung von anderen Arten von Aussagen bleibt eine Herausforderung.[166] Selbst wenn die Aussage einer Nachricht an sich korrekt ist, so bedeutet das nicht, dass die Daten nicht so ausgewählt bzw. in einen solchen Kontext gesetzt wurden, dass sie bewusst vage sind und missinterpretiert werden können.[178] Wu et al. versuchen dieses Problem automatisiert zu lösen, indem Aussagen als Parameter wahrgenommen werden. Werden diese Parameter anschließend verändert und in veränderter Form erneut geprüft, soll ihre ‚Robustheit' festgestellt werden. Eine schwache Robustheit weisen Aussagen auf, die nach einer kleinen Veränderung ihrer Parameter lediglich schwächere oder gegenteilige Schlussfolgerungen zulassen.[178] Die Aussagen einer Nachricht werden also in eine mathematische Funktion überführt.[179] Bei besonders komplex oder unüblich formulierten Aussagen kann dies eine Herausforderung sein.

Nahezu alle vollständig automatisierten Prüfungsansätze stoßen an ihre Grenzen, wenn externe Quellen zur Überprüfung der Korrektheit einer Aussage nicht verfügbar sind.[180]

[176] Vgl. Shi, B. und Weninger, T. (2016): Fact Checking in Heterogeneous Information Networks, S.101f.
[177] Vgl. Graves, L. (2018): Understanding the Promise and Limits, S.5
[178] Vgl. Wu, Y. et al. (2014): Toward computational fact-checking, S.589f.
[179] Vgl. Wu, Y. et al. (2014): iCheck, S.1063f.
[180] Vgl. Pérez-Rosas, V. et al. (2018): Automatic Detection of Fake News, S.3392

6.1.2 Überprüfung der Formulierung von Aussagen

Eine Untersuchung der Formulierung, ohne die Aussagen inhaltlich zu überprüfen, kann ebenfalls Aufschluss über den voraussichtlichen Wahrheitsgehalt einer Nachricht geben. Vor allem werden dabei stilistische, komplexitätsbezogene und psychologische Aspekte betrachtet. Sowohl von Menschen als auch maschinell verfasste Fake News weisen Besonderheiten auf, die die Erkennung erleichtern. Eine grundlegende Klassifikation, ob eine Nachricht von Menschen oder Maschinen verfasst wurde[181], kann auf Aussagen aufmerksam machen, die einer genaueren Betrachtung unterzogen werden sollten. Dies führt aber prinzipiell noch zu keinem Urteil darüber, ob es sich um Fake News oder wahre Nachrichten handelt.

In der Literatur existieren verschiedene Kategorien hinsichtlich auf die Formulierung bezogener Auswertungsmöglichkeiten: lexikalische (wortbezogene), syntaktische und bereichsbezogene (in Falle von Nachrichten auf den journalistischen Bereich bezogene)[182], komplexitätsbezogene und psychologische Faktoren.[183] In manchen Untersuchungen werden lexikalische, syntaktische und bereichsbezogene Kategorien nicht getrennt behandelt, sondern zusammengefasst als stilistische Faktoren bezeichnet.[183]

Die Untersuchung von **stilistischen** Merkmalen beschäftigt sich mit der Frage, ob beispielsweise bestimmte Wörter wie Negationen, informale Sprache und Fragewörter, ebenso wie weitere Merkmale (etwa Zitate oder Punktsetzung) verwendet werden.[183] Sprache, die auf Täuschungsversuche hinweist, oder auch das Maß an Subjektivität in einem Text können einbezogen werden.[184] Eine genauere Aufschlüsselung in lexikalische, syntaktische und bereichsbezogene Faktoren ist möglich. Die lexikalische Betrachtung kann auf dem Satz-, Wort- oder Zeichenlevel erfolgen.[182] Mit dem ‚Bag-of-Words'-Ansatz wird jedes Wort einzeln betrachtet.[185] Dabei wird beispielsweise die Häufigkeit des Vorkommens einzelner Begriffe analysiert. Mit dieser Betrachtung bleiben Zusammenhänge und der Kontext unerschlossen. Dabei können beispielsweise die Anzahl aller Wörter, die durchschnittliche Anzahl der Buchstaben pro Wort sowie die Anzahl langer oder unüblicher Wörter herangezogen werden. Die syntaktische Betrachtung von Sätzen

[181] Vgl. Zellers, R. et al. (2019): Defending Against Neural Fake News, S.6f.
[182] Vgl. Shu, K. et al. (2017): Fake News Detection on Social Media, S.5f.
[183] Vgl. Horne, B. und Adali, S. (2017): This Just In, S.3f.
[184] Vgl. Shu, K. et al. (2017): Fake News Detection on Social Media, S.7
[185] Vgl. Conroy, N. et al. (2015): Automatic Deception Detection, S.2f.

umfasst eine Messung der Häufigkeit von Funktionswörtern und eine Untersuchung der Zeichensetzung. Möglich ist auch der Einsatz von Deep-Syntax-Analysen, wobei die Struktur von Sätzen durch Syntaxbäume wiedergegeben wird.[185] Auf den journalistischen Bereich bezogene Merkmale können ebenfalls ausgewertet werden, so z. B. die Anzahl an Zitaten oder die Anzahl und Länge von Graphen.[182] Komplexere Maßnahmen können eine Kombination dieser Betrachtungsweisen beinhalten. Häufig in Forschungsvorhaben untersuchte stilistische Merkmale beinhalten die Anzahl an Silben, Wörtern oder Sätzen und die Anzahl an Wörtern pro Satz.[186] Diese Werte können prozentual[187] oder als Durchschnittswert[188] ausgewertet werden. In diesem Zusammenhang relevant können die Anteile an Wörtern bestimmter Länge oder die Anteile bestimmter Wörter (wie Pronomen, Artikel, die Verwendung grammatischer Personen oder Zeitformen) sein.[187] Redundanzen und Diversität, Generalisierungen oder die Verwendung emotionaler oder passiver Wörter können ebenso festgestellt werden.[188] Horne und Adali erweitern eine solche Kombination zahlreicher Merkmale wie verschiedene Werte zur Wortanzahl durch unter anderem auch die Anzahl an Schimpfwörtern, Slang-Worten, vollständig großgeschriebenen Wörtern, die Anzahl der Zitate und durch eine Betrachtung der Punktsetzung.[189]

Die **Komplexität** eines Textes kann auf dem Satz- und dem Wortlevel gemessen werden. Ein Satz wird von Horne und Adali dann als komplex eingeschätzt, wenn er viele Wörter enthält und der Syntaxbaum des Satzes tief ist.[183] Die Komplexität eines Wortes bemisst sich daran, wie einfach es zu lesen ist. Dies ist abhängig von der Anzahl der Silben.[183] Ebenso werden Wortwiederholungen und die Gebräuchlichkeit der verwendeten Wörter gemessen.[183]

Psychologische Aspekte des Schreibstils betreffen beispielsweise die Stimmung, die der Autor mit dem Text ausdrücken oder auslösen möchte. Häufig nutzen Fake-News-Autoren emotionsgeladene Sprache, insbesondere im Negativen, die beispielsweise anhand der Identifikation von Übertreibungen erkannt werden kann.[185]

Kombinationen solcher Aspekte können zu qualitativ unterschiedlichen Ergebnissen hinsichtlich der Genauigkeit der Klassifikation von Fake News führen. Ebenso können die Ergebnisse unterschiedlich ausfallen, je nachdem welche Bereiche (z. B. nur Textkörper

[186] Vgl. Burgoon, J. et al. (2003): Detecting Deception through Linguistic Analysis, S.93f.
[187] Vgl. Newman, M. et al. (2003): Lying words, S.670f.
[188] Vgl. Zhou, L. et al. (2004): Automating Linguistics-Based Cues for Detecting Deception, S.94f.
[189] Vgl. Horne, B. und Adali, S. (2017): This Just In, S.4ff.

oder Titel) analysiert werden.[189] Einige Aspekte können mittels einer Sentimentanalyse (Sentiment-Analysis) erhoben werden. Dabei werden wissensbasierte Ansätze (z. B. eine Klassifikation mittels vorgegebener, eindeutig positiver oder negativer Worte), statistische Ansätze (z. B. eine Analyse des Textes durch Machine Learning) oder Hybridansätze verwendet.[190] Das Ziel dabei ist es, automatisiert festzustellen, welche Meinung, Emotion oder welchen Grad an Subjektivität eine Nachricht aufweist.

Analysen von korrekten Nachrichten und Fake-News-Texten zeigen, dass legitime Nachrichtentexte in der Regel länger sind (und längere Wörter enthalten) als Fake-News-Texte.[189] Fake News beinhalten außerdem häufig weniger technische Wörter, Satzzeichen und Zitate, größere Negativität, weisen mehr Redundanz auf und sind insgesamt einfacher zu lesen.[189] Die Überschriften von Fake News enthalten hingegen mehr Inhalte als jene legitimer Nachrichten, die kürzer und allgemeiner sind.[189] Laut Aldwairi und Alwahedi ist in Überschriften von Fake News ebenso eine größere Menge an Satzzeichen zu erwarten als bei Überschriften wahrer Nachrichten.[191] Daher gilt laut den Autoren ab einer bestimmten Wortmenge in Überschriften oder in externen Links eine Nachricht als verdächtig hinsichtlich Fake News. Diese Vorgehensweise funktioniert vor allem bei Artikeln, die aufmerksamkeitserregend verfasst sind. Fake News, die hingegen als seriöse Pressemeldungen getarnt werden, orientieren sich an einem journalistischen Schreibstil und sind auf diese Weise schwieriger zu erkennen.

Die Einbeziehung des Inhaltes zur Feststellung, ob es sich um Fake News handelt, ist naheliegend. Dabei hat dieses Vorgehen sowohl Vor- als auch Nachteile. Vorteilhaft ist, dass eine inhaltliche Prüfung der Aussagen zuverlässige Informationen über den Wahrheitsgehalt bieten kann, da die jeweiligen Behauptungen direkt überprüft werden. Ebenso kann der verwendete Schreibstil vor allem bei reißerisch formulierten Nachrichten Hinweise auf das Vorhandensein von Fake News geben.

Es sind allerdings auch einige Nachteile zu beachten. So ist eine Faktenüberprüfung schwierig und aufwändig und mit der aktuell vorhandenen Software vor allem bei vage oder in komplexer Satzstruktur formulierten Aussagen schwierig automatisiert umzusetzen. Dies gilt ebenso für wenig erschlossene Bereiche, in denen Algorithmen auf keiner Datenbasis aufbauen können. In solchen Fällen eignet sich die automatisierte Faktenkontrolle als Unterstützung für menschliche Prüfer. Ebenso muss der verwendete Schreibstil

[190] Vgl. Cambria, E. et al. (2013): New Avenues in Opinion Mining and Sentiment Analysis, S.18f.
[191] Vgl. Aldwairi, M. und Alwahedi, A. (2018): Detecting Fake News in Social Media Networks, S.218

nicht zwingend Rückschlüsse auf den Wahrheitsgehalt einer Nachricht zulassen. Gelungene Fake News orientieren sich mitunter an einem professionellen journalistischen Schreibstil, der beispielsweise durch die Nutzung von Meldungen von Presseagenturen als Vorlage erreicht wird.[192] Eine Herausforderung ist auch die Unterscheidung von Fake News und stilistisch ähnlichen Formaten (beispielsweise Satire). Wird eine Sentimentanalyse eingesetzt, können vor allem bei simpleren Analysemethoden wie der einfachen Schlüsselworterkennung Fehler auftreten, wenn z. B. negative Schlüsselworte im Kontext positiv gemeint sind oder ein einzelner Satz die gesamte Bedeutung des vorhergehenden Abschnitts umkehrt.

6.2 Betrachtung der Quelle

Statt den Inhalt einer Nachricht zu analysieren, kann auch die Quelle betrachtet werden, die eine Nachricht in Umlauf gebracht hat. Damit in Verbindung steht auch die Reputation-Heuristic (siehe Kapitel 3.1). Die Glaubwürdigkeit einer Quelle spielt eine entscheidende Rolle dahingehend, ob eine Information von den Empfängern geglaubt wird oder nicht. Die Quelle einer Nachricht zu betrachten, um einzuschätzen, ob es sich um Fake News handelt, kann zusätzliche Hinweise liefern und bei einer großen Menge an Nachrichten von derselben Quelle die Erkennung beschleunigen. Die Grundannahme dieses Ansatzes ist es, dass eine Quelle von Nachrichten, z. B. eine Webseite oder ein Account in Social Media, die in der Vergangenheit Fake News verbreitet hat, dies vermutlich bei aktuellen und zukünftigen Nachrichten auch tun wird. Gegenteiliges gilt für als seriös einzuschätzende Quellen. Diese Methode wird bereits von Journalisten eingesetzt[193], könnte aber mittels Automatisierung schneller und kostengünstiger durchgeführt werden.

Innerhalb von Social Media ist die Quelle einer Nachricht ein Nutzeraccount bzw. -profil. Die von anderen Nutzern wahrgenommene Glaubwürdigkeit eines Social-Media-Accounts hängt von verschiedenen Faktoren ab, etwa vom Einfluss des Nutzers, der Anzahl der Follower, dem Ruf, dem Nutzernamen, der Expertise und den bisher veröffentlichten

[192] Vgl. Fraunhofer FKIE (2019): Software für die automatisierte Erkennung von Fake News
[193] Vgl. NewsGuard (2020): Bewertungsprozess und Kriterien

Nachrichten mit glaubwürdigen Links.[194] Ebenso relevant kann sein, wie lange ein Nutzer bereits auf einer Plattform angemeldet ist und wie viele Nachrichten er bisher verfasst hat.[195] Bei der Betrachtung des Nutzernamens ist nicht nur die Länge relevant[196], sondern auch die Bedeutung des Namens. Passt der Nutzername zum Inhalt der Nachricht, so wirkt dies glaubwürdiger.[198] Solche Informationen können auch im Rahmen einer Automatisierung herangezogen werden. Steht hinter einem Account ein Social Bot, kann das ebenfalls einen Hinweis auf mögliche Fake-News-Aktivitäten geben. Menschen versuchen andere Nutzer von Bots zu unterscheiden, indem sie die Beschreibung des Profils, den Profilnamen, die Anzahl der Follower, hochgeladene Bilder und die Historie der vom Nutzer geschriebenen Nachrichten einbeziehen.[197] Ist ein Profilbild, das standardmäßig voreingestellt ist, nicht verändert worden, mindert dies beispielsweise die wahrgenommene Glaubwürdigkeit.[198] Automatisierte Erkennungsalgorithmen ziehen teilweise andere Merkmale heran. Unter Einbeziehung von Accountinformationen wie Entropie (der Pause zwischen Nachrichtenveröffentlichungen), URL-Ratio (wie häufig in allen Nachrichten eine externe URL beinhaltet ist) und den Geräten, mit denen die Nachrichten veröffentlicht werden, erreichen Chu et al. eine durchschnittliche Klassifikationsgenauigkeit in die Kategorien Mensch, Bot und Cyborg von 96 %.[199] Auch Davis et al. haben mit ‚BotOrNot' ein System entwickelt, das auf Basis von inhaltlichen, netzwerkbasierten und nutzerbasierten Merkmalen bewertet, ob es sich bei einem Nutzer eher um einen Menschen oder einen Social Bot handelt.[200] Hilfreich bei einer automatisierten Überprüfung von Account-Eigenschaften können für die Erkennung von einfachen Bot-Accounts auch die Merkmale der numerischen ID (da falsche Accounts häufig zeitlich nah beieinander erstellt werden) und die Ratio zwischen Followern und den Accounts sein, denen ein vermeintlicher Bot folgt.[194] Für die Auswertung ebenso relevant zeigen sich in anderen Untersuchungen die Anzahl an Status-Updates eines Accounts, die Frage, ob ein Account einen URL-Link aufweist, und die Nutzung von Personalpronomen.[197] Ein persönlicher Schreibstil soll Hinweise auf einen menschlichen Autor geben. Die Information, ob es sich bei einer Quelle um einen Social Bot oder einen Menschen handelt, könnte dann beispielsweise von anderen Systemen genutzt werden, um mögliche Verursacher von

[194] Vgl. Appling, S. und Briscoe, E. (2017): The Perception of Social Bots, S.21
[195] Vgl. Castillo, C. et al. (2011): Information credibility on twitter, S.680
[196] Vgl. Gupta, A. et al. (2014): TweetCred, S.228f.
[197] Vgl. Appling, S. und Briscoe, E. (2017): The Perception of Social Bots, S.23f.
[198] Vgl. Morris, M. et al. (2012): Tweeting is Believing, S.445
[199] Vgl. Chu, Z. et al. (2012): Detecting Automation of Twitter Accounts, S.811f.
[200] Vgl. Davis, C. et al. (2016): BotOrNot, S.273f.

Fake News zu identifizieren. In Botnetzen vernetzte oder generell komplexere Social Bots sind schwieriger zu identifizieren, da sie zu einem höheren Grad menschliches Verhalten simulieren.

Die Quelle beeinflusst auch die Informationen hinsichtlich der Veröffentlichung einer Nachricht. Solche Metadaten, z. B. zur Frage, vor wie vielen Sekunden eine Nachricht veröffentlicht wurde oder ob sie über Koordinaten verfügt, können ebenfalls zur Beurteilung des Wahrheitsgehalts einer Nachricht herangezogen werden.[201] Diese Informationen sind teilweise für menschliche Leser nicht oder lediglich mit Mühe aufzudecken, können aber über Schnittstellen mithilfe von Software einfacher abgerufen werden.[202]

Die Betrachtung der Quelle kann somit Hinweise auf das Vorhandensein von Fake News liefern. Dies hat sowohl Vor- als auch Nachteile. Besonders bei bekannten oder offensichtlichen Quellen von Fake News kann eine quellenbezogene Vorsortierung während der Fake-News-Prüfung arbeitserleichternd wirken. Weniger komplexe Fake-News-Accounts und Social Bots können mit automatisierten Mitteln identifiziert werden.

Mögliche Nachteile zeigen sich vor allem bei komplexeren Social Bots, die durch ihre Tarnung als Menschen mit teilweise korrekten Meldungen und unauffälligem Verhalten schwierig zu erkennen sind. Ein weiteres mögliches Problem offenbart sich bei der Vorverurteilung von Quellen. Eine unvollständige Betrachtung einer Quelle kann nicht immer ausreichend sein, um zu beurteilen, ob es sich um Fake News handelt, da auch Fake-News-Quellen korrekte Nachrichten veröffentlichen oder legitime Quellen Fehler machen können.[203]

[201] Vgl. Gupta, A. et al. (2014): TweetCred, S.233
[202] Vgl. Shariff, S. et al. (2017): On the credibility perception of news on Twitter, S.794
[203] Vgl. Graves, L. (2018): Understanding the Promise and Limits, S.7

6.3 Betrachtung der Umgebung

Die Umgebung, in der sich Nachrichten befinden, kann ebenfalls Aufschluss über den Wahrheitsgehalt geben. Dieser Ansatz betrachtet die Eigenschaften und das Verhalten des die Nachricht umgebenden Netzwerks und somit auch dessen Reaktionen darauf.[204] Es kann eine Betrachtung der Nutzer im Umfeld und eine Betrachtung der Reaktionen von Nutzern auf eine Nachricht vorgenommen werden.[205]

Das **Netzwerk** zu untersuchen, in dem sich die Quelle befindet, ist ein Teilansatz dieser Betrachtungsweise, um abzuleiten, ob es sich bei einer Nachricht um Fake News handelt.[205] Dabei wird die Annahme getroffen, dass die Quellen von korrekten Nachrichten und von Fake News jeweils unterschiedliche Netzwerke aufbauen, die verschiedene Charakteristika aufweisen, z.B. die durchschnittliche Anzahl an Freunden oder Followern und die Anzahl verifizierter Nutzer. Ein solches Netzwerk kann homogen mit zahlreichen gleichen Akteuren oder heterogen mit unterschiedlichen Akteuren aufgebaut sein.[205] Gupta et al. nutzen in diesem Zusammenhang einen Ansatz, bei dem davon ausgegangen wird, dass Nutzer mit einem hohen Glaubwürdigkeitswert wahrscheinlicher mit einer höheren Anzahl an glaubwürdigen Nutzern vernetzt sind als mit unglaubwürdigen Nutzern.[206] Das Gleiche wird für Nachrichten in Social Media angenommen. Dieses Maß an Glaubwürdigkeit wird zu einem gewissen Maß an die direkt vernetzten Nachbarn übertragen. Dadurch ergibt sich die Glaubwürdigkeit einer Nachricht nicht nur aus ihrem direkten Umfeld, sondern auch indirekt aus dem Umfeld ihres Umfelds. Ebenso kann davon ausgegangen werden, dass verschiedene Fake-News-Seiten oder -Accounts mehr gemeinsame Nutzer aufweisen als mit Accounts anderer Art.[207] Dadurch bilden sich Communities, über die ebenfalls Rückschlüsse auf den Wahrheitsgehalt der darin befindlichen Nachrichten möglich sind. Wenn Nachrichten lediglich von einer geringen Anzahl an Nutzern im Netzwerk stammen und häufig weiterverbreitet werden, ist dies nach Erkenntnissen von Castillo et al. ein Indikator für eine höhere Glaubwürdigkeit von Nachrichten.[208]

[204] Vgl. Conroy, N. et al. (2015): Automatic Deception Detection, S.3f.
[205] Vgl. Shu, K. et al. (2017): Fake News Detection on Social Media, S.5ff.
[206] Vgl. Gupta, M. et al. (2012): Evaluating Event Credibility on Twitter, S.156
[207] Vgl. Tacchini, E. et al. (2017): Some Like it Hoax, S.4
[208] Vgl. Castillo, C. et al. (2011): Information credibility on twitter, S.682

Eine Betrachtung der Durchdringung eines Netzwerks mit Social Bots, sofern diese erfolgreich erkannt werden können, kann ebenso Hinweise geben. Werden die Nutzer im Netzwerk der Quelle auf ihre Menschlichkeit oder Glaubwürdigkeit hin analysiert, können ähnliche Verfahren wie bei der Betrachtung der eigentlichen Quelle eingesetzt werden, um beispielsweise die Glaubwürdigkeit aller Nutzer im Umfeld individuell oder im Durchschnitt als Gruppe zu bewerten.[205]

Die **Reaktion anderer Nutzer** im Umfeld der Quelle auf eine Nachricht kann unterschiedlich bemessen werden. Vor allem die Reaktion menschlicher Nutzer auf eine Nachricht kann deutliche Hinweise geben und bildet damit einen sozialen Kontext.[209] Auch als Haltung der Nutzer bezeichnet[205], werden die Meinungen zur Nachricht (z. B. Zustimmung oder Ablehnung) betrachtet. Diese Reaktion, die die Haltung eines Nutzers auf eine Nachricht widerspiegelt, bildet eine Beziehung zwischen der veröffentlichenden Quelle und dem auf die Nachricht reagierenden Nutzer und kann von Unterstützung oder Ablehnung geprägt sein. Mehrere Reaktionen auf eine Nachricht bilden damit ein Netzwerk aus Unterstützern und Gegnern.[210] Die Haltung eines Nutzers kann explizit oder implizit vorliegen.[209] Explizite Haltungen zeigen sich als unmittelbarer Ausdruck einer Reaktion oder Emotion, z. B. mit ‚Daumen hoch' auf Facebook. Implizite Haltungen ergeben sich z. B. aus Nachrichten eines Nutzers. Um solche Informationen automatisiert zu erheben, kann eine Sentimentanalyse durchgeführt werden (siehe Kapitel 6.1.2). Die Auswertung der Reaktionen erfolgt dabei unbewusst, d. h., die Nutzer werden in der Regel nicht darüber informiert, dass aus ihrer Reaktion der Wahrheitsgehalt einer Nachricht abgeleitet werden soll. Dies kann dabei helfen, Ergebnisse nicht zu verfälschen, und reduziert den Aufwand, da es keiner Vorbereitung der Nutzer bedarf. Mittels semantischer Analysen können die Aussagen eines Nutzers anschließend mit jenen anderer Nutzer abgeglichen werden.[211] Sagen mehrere Nutzer (z. B. bei einem Erlebnisbericht) das Gleiche aus, wird davon ausgegangen, dass dies der Wahrheit entspricht.

Der Ansatz, das Umfeld einer Nachricht zu untersuchen, weist ebenso wie die anderen Herangehensweisen Vor- und Nachteile auf. Vorteilhaft ist das Einbeziehen zahlreicher anderer Nutzer, um Fake News zu erkennen. Somit kann kostenlos auf eine erste menschliche Vorbeurteilung zurückgegriffen werden. Durch eine größere Stichprobe kann im

[209] Vgl. Shu, K. et al. (2017): Fake News Detection on Social Media, S.7
[210] Vgl. Jin, Z. et al. (2016): News Verification by Exploiting Conflicting Social Viewpoints, S.2972
[211] Vgl. Conroy, N. et al. (2015): Automatic Deception Detection, S.2f.

Sinne von Schwarmintelligenz versucht werden, inhaltlich vage formulierte oder anderweitig automatisiert schwierig zu erkennende Fake News aufzudecken.

Von Nachteil ist hingegen, dass dieser Ansatz lediglich funktioniert, wenn eine Nachricht bereits von genügend Nutzern gesehen und darauf reagiert wurde bzw. ausreichend Nutzer im Netzwerk zur Verfügung stehen und Inhalte produzieren. Daher kann eine Verzögerung bei der Auswertung nötig sein, um hinreichende Ergebnisse zu erzielen und ausreichend Daten zu generieren.[212] Handelt es sich um schädliche Fake News, sind die Nachrichten damit bereits in Umlauf geraten und einigen Nutzern bekannt, bei denen sie ihre schädliche Wirkung entfalten konnten. Außerdem besteht jederzeit die Gefahr, dass besonders aufwändig hergestellte Fake News von einem Großteil der Nutzer nicht als solche erkannt werden und somit die Reaktionen fälschlicherweise auf korrekte Nachrichten schließen lassen. Dies würde auch Bewertungsverfahren, in denen mehrere Nutzer über den Wahrheitsgehalt einer Nachricht abstimmen sollen, verfälschen. Ebenso können bei der Erkennung der Haltung eines Nutzers, z. B. im Rahmen einer Sentimentanalyse, bei nicht geradeheraus formulierten Reaktionen Fehler während der Klassifikation auftreten und somit falsche Rückschlüsse gezogen werden. In diesem Zusammenhang relevant ist auch die Unterscheidung von menschlichen Nutzern und Social Bots eines Botnetzwerks, die beispielsweise darauf programmiert werden können, zu Fake News Zustimmung auszudrücken. Werden bei der automatisierten Auswertung der Umgebung und deren Reaktionen Social Bots nicht erkannt, wird das Ergebnis verfälscht.

[212] Vgl. Jin, Z. et al. (2016): News Verification by Exploiting Conflicting Social Viewpoints, S.2977

6.4 Kombinationen verschiedener Ansätze

Lediglich einen der genannten Ansätze oder einen Teilbereich davon einzusetzen (z. B. eine reine Analyse des Schreibstils), kann zu ungenauen Ergebnissen führen, die keine hinreichende Klassifikation ermöglichen.[213] Um eine möglichst korrekte Einschätzung einer Nachricht in Bezug auf mögliche Fake News zu erhalten und bei einer Automatisierung flexibel auf veränderliche Informationsangebote reagieren zu können, ist eine Kombination der oben beschriebenen Ansätze vielversprechend. In einigen Forschungsvorhaben werden bereits kombinierte Ansätze verwendet. Je nach Bedarf können verschiedene Merkmale einer Nachricht erhoben werden. In einem Forschungsvorhaben am Fraunhofer Institut wurden alle drei Ansätze kombiniert, indem der Schreibstil der Nachricht neben Metadaten zur Quelle (z. B. zur Frage, wie oft ein Nutzer Nachrichten veröffentlicht) oder zum Netzwerk betrachtet wurde.[214] Die Auswertungsmöglichkeiten hängen dabei von den Merkmalen einer Nachricht ab. Sollte beispielsweise der Text einer Nachricht für eine inhaltliche Analyse nicht lang genug sein, kann die Betrachtung anderer Merkmale in den Vordergrund treten, z. B. die Analyse der vorhandenen Links oder der Quelle und ihres Netzwerks.[215] Auch Della Vedova et al. setzen eine Kombination aus inhaltlicher und reaktionsbasierter Betrachtung auf Basis der Menge der vorhandenen Informationen ein.[216] Solange die Anzahl der Reaktionen einen gewissen Schwellenwert unterschreitet, wird eine inhaltsbasierte Klassifikation in Fake News und wahre Nachrichten durchgeführt. Sind ausreichend soziale Signale in Form von Reaktionen anderer Nutzer vorhanden, wird eine Klassifikation auf Basis dieser durchgeführt. Ruchansky et al. nutzen für ihr CSI- Modell (Capture, Score und Integrate) eine Kombination aller Ansätze.[217] Mittels eines neuralen Netzwerks werden zunächst notwendige Informationen wie das Verhalten der Nutzer gesammelt, anschließend analysiert und ein Artikel abschließend als falsch oder wahr klassifiziert.[218] Der Text der Nachricht, die Reaktion anderer Nutzer und die Betrachtung der veröffentlichenden Quelle werden als Informationen ebenso herangezogen.[218]

[213] Vgl. Potthast, M. et al. (2018): A Stylometric Inquiry into Hyperpartisan and Fake News, S.237f.
[214] Vgl. Fraunhofer FKIE (2019): Software für die automatisierte Erkennung von Fake News
[215] Vgl. Hen, J. (2020): Automatisierte Wahrheitssuche
[216] Vgl. Della Vedova, M. et al. (2018): Automatic Online Fake News Detection, S.274ff.
[217] Vgl. Ruchansky, N. et al. (2017): CSI, S.802
[218] Vgl. Ruchansky, N. et al. (2017): CSI, S.797f.

Da nicht davon auszugehen ist, dass bei den zu untersuchenden Nachrichten stets alle notwendigen Informationen vorliegen, eignet sich zur Erkennung von Fake News vor allem ein kombinierter Ansatz. Vorteilhaft ist hierbei, dass die Schwächen einzelner Methoden nicht mehr ins Gewicht fallen, da sie durch die Stärken anderer ausgeglichen werden können. Es werden zahlreiche Informationen erhoben, die andere, fehlende Merkmale ersetzen können und das Modell flexibler machen. Gleichzeitig kann sich durch die größere Menge an betrachteten Merkmalen die Klassifikationsgenauigkeit erhöhen.

Ein Nachteil kann sein, dass der Aufwand, ein solches System aufzubauen, umso größer ist, je mehr Merkmale erhoben und analysiert werden sollen.

7 Aktuelle Herausforderungen im Umgang mit Fake News

Die Erkennung von Fake News ist eine bedeutende Aufgabe für deren Bekämpfung. In diversen Forschungsvorhaben konnten dazu bisher Fortschritte erzielt werden. Trotzdem ergeben sich weiterhin einige Herausforderungen und Probleme, die im Zuge weiterer Forschungsvorhaben aufgegriffen werden können.

Aktuell bleibt eine rein automatisierte Überprüfung von Nachrichten in der Regel auf spezifische, eng gefasste Bereiche beschränkt, häufig auch auf Aussagen mit statistischen Behauptungen, die einfach überprüft werden können.[219] Vage Aussagen oder Halbwahrheiten erschweren die korrekte Erkennung von Fake News. Die Leistung von Systemen zur Identifikation von Falschnachrichten kann dabei bereichsspezifisch sein und bezüglich der Themenbereiche (z. B. Politik, Unterhaltung) oder der untersuchten Sprachen variieren.[220] Dadurch ist ein übergreifendes Erkennungssystem für alle Sprachen und Themenbereiche aktuell schwer realisierbar. Sofern Nachrichten nicht direkt als Text, sondern in anderer Form (z. B. als Video oder Bild) vorliegen, wird die Erkennung von Fake News dadurch erschwert, dass kein direkt verarbeitbarer Text vorliegt. Wenn auf eine automatisierte inhaltliche oder sprachliche Analyse nicht verzichtet werden kann, müssen weitere Technologien, z. B. Spracherkennung, miteinbezogen werden. Damit können zusätzliche Fehlerquellen in die Untersuchung gebracht werden.

Je nachdem wie streng ein System bewertet, besteht die Gefahr, dass entweder zahlreiche wahre Nachrichten als Fake News eingeordnet werden (False Positives) oder Fake News als wahre Nachrichten durchgehen (False Negatives). Eine vollständig korrekte Einordnung ist nicht zu erreichen, trotzdem sollte die Genauigkeit möglichst hoch sein. Für schwierig einzuordnende Grenzfälle oder bei Themen, die zahlreiche Nuancen beinhalten, kann daher aktuell das Einbeziehen menschlicher Bewerter (z. B. mit einem automatisierten System als Vorbetrachter) notwendig bleiben. Fälle, in denen menschliche Unterstützung notwendig sein könnte, müssten jedoch selbstständig vom System erkannt werden.

[219] Vgl. Graves, L. (2018): Understanding the Promise and Limits, S.5
[220] Vgl. Pérez-Rosas, V. et al. (2018): Automatic Detection of Fake News, S.3398

Die automatisierte Erstellung und Verbreitung von Fake News und das Nachvollziehen der dafür verwendeten Mechanismen sind ebenfalls nicht vollständig untersucht. Dies kann damit begründet werden, dass Ersteller von Fake News kein Interesse daran haben, ihre Methoden aufzudecken, da dadurch die Erkennung erschwert wird. Umso besser aktuell eingesetzte Erstellungs- und Verbreitungsmethoden verstanden werden, desto effektiver können diese gezielt bekämpft werden.

Eine Herausforderung beim Training von Algorithmen bleibt der Mangel an einer ausreichend großen, für die Zwecke des maschinellen Lernens geeigneten Datenbasis.[219] Vor allem bei nicht populären Themen kann dies ein Problem sein. Fake-News-Beispiele aus der Realität, die nicht konstruiert wurden, vermeiden mögliche Fehler bei der Erkennung, indem tatsächlich im Umlauf befindliche und veröffentlichte Fake News einbezogen werden. Selbst konstruierte Fake News können stets lediglich den aktuellen Stand des Wissens über Fake News abbilden. Die Analyse tatsächlicher Fake News kann diesen Wissensstand erweitern und somit auch die Erkennung verbessern. Dabei ist es für eine noch folgende erfolgreiche Untersuchung ausschlaggebend, ausreichend Beispiele sowohl für Fake News als auch für legitime Nachrichten zum angestrebten Themengebiet zu finden und dabei keine False Positives oder False Negatives im Datensatz zuzulassen. Aufgrund aktuell noch bestehender Uneinheitlichkeit hinsichtlich der Definition von Fake News können aufgrund unterschiedlicher Auffassungen hinsichtlich der Frage, bei welchen Nachrichten es sich um Fake News handelt, verschiedene Systeme zu unterschiedlichen Ergebnissen kommen. Eine allgemein akzeptierte, einheitliche Definition inklusive einer Abgrenzung von ähnlichen Konzepten (z. B. Satire) fehlt bislang.

Für die automatisierte Erkennung von Fake News existieren bereits verschiedene Ansätze, die unterschiedliche Merkmale erheben und damit zu unterschiedlichen Ergebnissen kommen können. Eine reine inhaltliche Betrachtung würde theoretisch ausreichen, um über Wahrheit oder Fälschung zu entscheiden, da lediglich auf diese Weise der Inhalt einer Nachricht direkt einbezogen wird. Dabei müssen auch komplex formulierte Nachrichten automatisiert nahezu vollständig korrekt verstanden und zeitnah verarbeitet werden können. Da dies zurzeit nicht der Fall ist und dies zum Teil auch von Menschen nicht erreicht werden kann, eignet sich die Einbeziehung weiterer Ansätze, etwa einer Betrachtung der Quelle und der Umgebung, um mittels weiterer Merkmale Fake News sicher und schnell zu erkennen. Bisher gibt es keine allumfassende Untersuchung, mit der herausge-

stellt wurde, welche Merkmale sowohl innerhalb der Ansätze als auch welche Kombinationen von Ansätzen für welchen Einsatzbereich zur Erkennung von Fake News das beste Endergebnis liefern. Das Einbeziehen von möglichst vielen Merkmalen kann dabei für adäquate Ergebnisse sorgen, aber die Erkennung auch unnötig verlangsamen und Ressourcen beanspruchen, da nicht bekannt ist, welche Merkmale welchen Einfluss auf die Güte des Klassifikationsergebnisses haben. Eine Gegenüberstellung bisheriger Ergebnisse und weitere die Effektivität der relevanten Merkmale betrachtende Untersuchungen können zukünftige Forschungen vereinfachen und verbessern. Dabei sollte allerdings die Entwicklung der Fake-News-Erstellung, -Verbreitung und das Umfeld beobachtet und bei Veränderungen bei Bedarf darauf reagiert werden.

Da Fake News bereits ab ihrem ersten erreichten Empfänger Schaden anrichten können, sollten sie möglichst schnell identifiziert werden, damit gegen sie vorgegangen werden kann. Dabei ergibt sich die Problematik, dass (abgesehen von rein inhaltlichen Informationen) erst eine gewisse Anzahl von Interaktionen seitens der Quelle oder seitens des Umfelds notwendig ist, um Rückschlüsse ziehen zu können. Die große Menge an Nachrichten, die in Social Media verschickt werden, stellt in dem Zusammenhang ebenfalls eine Herausforderung dar. Um eine solche Menge möglichst zeitnah überprüfen zu können, wären leistungsstarke Systeme nötig, die eine ausreichende finanzielle und personelle Unterstützung und ebenso eine passende Infrastruktur benötigen. Ein solches großes, zentralisiertes Projekt existiert (in Deutschland) zu diesem Zeitpunkt nicht. Mehrere kleine Teilprojekte, die z. B. bedeutende Themenbereiche abdecken, könnten relevante Teilarbeit leisten. Dabei ergibt sich aber die Frage nach Überschneidungen, da es ineffizient wäre, wenn beispielsweise mehrere Systeme dieselbe Nachricht überprüfen. Eine engere Vernetzung und ein Informationsaustausch zwischen diesen Projekten, auch international, könnten für eine effizientere Bekämpfung sorgen. Auch Social-Media-Plattformen, auf denen Fake News veröffentlicht werden, können aktuell lediglich unzureichend in Projekte dieser Art eingebunden werden. Vor allem im Social-Media-Bereich geben die beteiligten Plattformen an, gegen Fake News vorgehen zu wollen. Da dabei der Öffentlichkeit aber in der Regel keine konkreten Informationen zugänglich gemacht werden, kann keine Kontrolle durch unabhängige Außenstellen erfolgen, wodurch mögliche kooperative Forschungsvorhaben erschwert werden.

Die Erkennung von Fake News bildet einen Grundpfeiler der Bekämpfung von Fake News und deren Wirkung. Die Informierung der möglichen Empfänger kann diese Nutzer

für die Wirkmechanismen unempfänglicher machen. Da sich zahlreiche Menschen über Fake News unzureichend aufgeklärt fühlen, könnte Bildung darüber Sicherheit geben. Sobald ein Empfänger Fake News beispielsweise selbst erkennt und darauf entsprechend reagiert, kann dies im Gegenzug automatisierte Erkennungsvorhaben (z. B. bei der Betrachtung der Reaktion der Umwelt) unterstützen und effektiver machen. Passende Bildungsvorhaben sind vor allem auch im Social-Media-Bereich bisher unzureichend untersucht worden.

8 Kritische Würdigung

Diese Arbeit eignet sich aufgrund ihres begrenzten Umfangs vor allem als erste, grobe Übersicht und als Einstieg in das Thema ‚Fake News', vor allem im Social-Media-Kontext. Dabei erfolgte eine Betrachtung des aktuellen Standes von Literatur und Forschung, neue Lösungen zu Wirkmechanismen, zur Verbreitung und zur Erkennung wurden nicht angeboten. Da der Umfang und die Zeit, die dieser Arbeit gewidmet werden konnten, begrenzt waren, konnten nicht alle Quellen thematisiert und tiefgehend betrachtet werden. Detailinformationen wie die genauen Abläufe von Algorithmen und eine Gegenüberstellung von Kennzahlen wurden aus diesem Grund nicht behandelt, auch konnten Systeme zur Fake-News-Erkennung nicht selbst erneut implementiert und getestet werden. Dadurch standen lediglich die Informationen der zitierten Autoren und Studien zur Verfügung, die mitunter detaillierter hätten sein können.

Einige Aspekte von Fake News sind bisher noch weitgehend unerforscht, sodass dahingehend weniger Informationen zur Verfügung stehen. Dies ist beispielsweise bei der Untersuchung der Eigenschaften, die Personen für die Wirkmechanismen von Fake News anfällig machen können, der Fall. Auch im Bereich der Fake-News-Erkennung existieren bisher keine eindeutigen Aussagen darüber, welche Merkmale eine optimale Kombination bilden. Daher konnte an dieser Stelle lediglich ein Überblick über verschiedene Ansätze anstatt einer finalen Lösung angeboten werden.

Statt der eingehenderen Betrachtung weniger Forschungsvorhaben wurde eine breitere Übersicht über eine größere Menge an Systemen vorgenommen, um einen Überblick über den aktuellen Stand der bekannten Wirkmechanismen sowie der Möglichkeiten der Erstellung und Verbreitung und der Erkennung von Fake News zu bieten. Dafür wurde eine große Anzahl an Quellen einbezogen, um verschiedene Vorhaben in der Forschung zu systematisieren und in einen gemeinsamen Kontext zu setzen. Die Auswahl blieb dabei auf deutsch- und englischsprachige Publikationen beschränkt. Abgesehen von den vorgestellten Wirkmechanismen, Verfahren zur Verbreitung und Erkennungsansätzen existieren vermutlich weitere, die an dieser Stelle nicht genauer untersucht wurden. Mit weiteren Informationen aus diesem Bereich ist eine genauere Systematisierung denkbar, die hier nicht erreicht werden konnte. Diese Arbeit dient somit vor allem als Einstieg und Anknüpfungspunkt, von dem aus ein Leser einen Überblick und ein Grundverständnis über die Thematik erhält und anschließend in Teilbereichen tiefere Recherchen anstellen kann.

Der Bereich ‚Fake News' bleibt in der Literatur uneindeutig definiert und erhielt in dieser Arbeit auch keine allumfassende, abschließende Definition. Ebenso wurden keine Handlungsempfehlungen zur optimalen Bekämpfung von Fake News oder weitere Lösungs- bzw. Präventivvorschläge gemacht. Während die Einführungs-, Definitions- und Wirkmechanismen-Abschnitte aufgrund der Gegebenheiten von Fake News auch in anderen Bereichen angewandt werden könnten, beschränkten sich die darauffolgenden Abschnitte hauptsächlich auf den Social-Media-Bereich und die Automatisierung. Fake News treten nicht nur im Social-Media-Bereich auf, sondern sind in zahlreichen weiteren Lebensbereichen anzutreffen. Diese (z. B. andere Bereiche des Internets wie Webseiten oder Online-Videos) könnten bedeutende Untersuchungsthemen bieten, die hier nicht genauer betrachtet wurden. Weitere Ansatzmöglichkeiten für Forschungs- und Recherchevorhaben, die in dieser Arbeit aufgrund des Umfangs nicht genauer adressiert werden konnten, wurden mittels einer Betrachtung der aktuellen Herausforderungen im Umgang mit Fake News aufgezeigt. Vor allem der Vorgang der automatisierten Erstellung von Fake News bzw. die dazugehörigen Verbreitungsmethoden bleiben undurchsichtig, da den Verursachern häufig nicht daran gelegen sein dürfte, ihre Methoden offenzulegen. Daher konnte im Rahmen dieser Arbeit nicht tiefgreifend auf diese eingegangen werden. Eine großflächige Untersuchung auf Social Media, mit der die Entstehung von Fake News besser verstanden werden könnte, war aufgrund des begrenzten Umfangs dieser Arbeit nicht möglich.

Insgesamt ist dieser Text geeignet, dem Leser einen Überblick und einen Einstieg in die Wirkmechanismen, die Erstellung und die Erkennung von Fake News insbesondere im Social-Media-Bereich bereitzustellen, eignet sich aber nicht zur detaillierten Betrachtung einzelner Vorhaben.

9 Fazit

Das Ziel dieser Arbeit ist es, den aktuellen Stand der Forschung zu Wirkmechanismen, Verbreitung und Erkennung von Fake News vor allem im Social-Media-Bereich auf Basis einer umfassenden Literaturrecherche systematisch wiederzugeben. Es zeigte sich, dass zahlreiche Mechanismen existieren, die im Zusammenhang mit Fake News wirken können. Die Stärke der Wirkung ist dabei stets von den individuellen Eigenschaften des Empfängers abhängig. Die Wirkmechanismen nehmen vor allem dann Einfluss, wenn Fake News verbreitet werden und auf ein passendes Publikum treffen, auch im Social-Media-Bereich. Durch die besonderen Gegebenheiten des Internets im Allgemeinen und Social Media im Besonderen existieren für die Ersteller von Fake News effektive Verbreitungsstrategien. Häufig wird sich dabei Social Bots bedient, die sich, teilweise in Netzwerken zusammengeschlossen, als menschliche Nutzer ausgeben. Davon werden andere Nutzer getäuscht. Zahlreiche potenzielle Empfänger befinden sich auch in Deutschland und fühlen sich unzureichend über Fake News aufgeklärt. Mittels einer automatisierten Erkennung können menschliche Empfänger und Prüfer entlastet werden. Eine Untersuchung, wie Fake News automatisiert erstellt und verbreitet werden, kann dabei helfen, diese zu bekämpfen, indem wertvolle Informationen für deren Erkennung gewonnen werden. Hinsichtlich der Identifikation von Fake News existieren drei Ansätze, die sich jeweils auf den Inhalt einer Nachricht, die Quelle und die Umgebung konzentrieren. Kombinationen davon können ebenfalls zum Einsatz kommen, um die Schwächen der einzelnen Ansätze zu kompensieren. Dabei bleibt Potenzial für weitere Forschungsvorhaben und Kooperationen aufgrund möglicher Weiterentwicklungen bezüglich der Erstellung von Fake News, bisher bestehender Heterogenität bei der Erkennung und unzureichend erforschter Bereiche. Ein koordiniertes Vorgehen gegen Fake News kann von Vorteil sein. Es ist davon auszugehen, dass in Zukunft weiterhin Fake News existieren werden, die unterschiedliche und aktuell relevante Themen aufgreifen. Aus diesem Grund kann eine anhaltende Untersuchung der Wirkmechanismen und der Methoden der Verbreitung und Erkennung von Fake News auch zukünftig mögliche Gegenmaßnahmen hervorbringen. Dadurch kann Fehlinformation von den Empfängern von Fake News abgewendet werden. Weitere Forschungsvorhaben eignen sich dazu, verwendete Verbreitungsmethoden genauer zu untersuchen, aufzudecken und auf Basis der gewonnenen Erkenntnisse beste-

hende Erkennungsmethoden zu verbessern. Ebenso kann angestrebt werden, neue Erkennungsmethoden und -ansätze zu gewinnen, um die Wirkung von Fake News und damit einhergehende potenzielle Schäden auf den Empfänger flächendeckend zu minimieren. Eine automatisierte Erkennung in diesem Bereich kann eingesetzt werden, um Nutzer im Social-Media-Bereich und auch im restlichen Internet zu schützen.

IV Literaturverzeichnis

Abdul-Kader, S.; Woods, D. (2015*). Survey on chatbot design techniques in speech conversation systems*. International Journal of Advanced Computer Science and Applications, 6(7), S.72–80. Science and Information Organization, West Yorkshire.

Abokhodair, N.; Yoo, D.; McDonald, D. (2015). *Dissecting a Social Botnet*: *Growth, Content and Influence in Twitter.* In: Cosley, D.; Forte, A.; Ciolfi, L.; McDonald, D. (Hg.): Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15, S.839–851. ACM Press, New York.

Adair, B.; Li, C.; Yang, J.; Yu, C. (2017). *Progress Toward "the Holy Grail": The Continued Quest to Automate Fact-Checking.* Computation + Journalism Symposium, Evanston.

Akademische Gesellschaft (2018). *How powerful are Social Bots?* Online im Internet: http://www.akademische-gesellschaft.com/fileadmin/webcontent/Publikationen/Communication_Snapshots/AGUK_CommunicationSnapshot_Social Bots_June2018.pdf (Abrufdatum: 24.09.20), veröffentlicht im Juni 2018.

Aldwairi, M.; Alwahedi, A. (2018). *Detecting Fake News in Social Media Networks*. Procedia Computer Science 141, S.215-222. Elsevier, Amsterdam.

Allcott, H.; Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. Journal of Economic Perspectives 31 (2), S.211-236. American Economic Association, Nashville.

Andrews, E. (2019). *How fake news spreads like a real virus*. Online im Internet: https://engineering.stanford.edu/magazine/article/how-fake-news-spreads-real-virus (Abrufdatum: 22.09.20), veröffentlicht am 09.10.19.

Appling, S.; Briscoe, E. (2017). *The Perception of Social Bots by Human and Machine*. Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference, S.20-25. Association for the Advancement of Artificial Intelligence, Menlo Park.

ARD; ZDF (2019). *Nutzung von Onlinecommunitys 2019.* Online im Internet: http://www.ard-zdf-onlinestudie.de/whatsapponlinecommunities/ (Abrufdatum: 21.09.20), veröffentlicht im Oktober 2019.

Arik, S.; Chrzanowski, M.; Coates, A.; Diamos, G.; Gibiansky, A.; Kang, Y.; Li, X.; Miller, J.; Ng, A.; Raiman, J.; Sengupta, S.; Shoeybi, M. (2017). *Deep Voice: Real-time Neural Text-to-Speech.* Online im Internet: http://ar xiv.org/pdf/1702.07825v2 (Abrufdatum: 29.09.20), veröffentlicht am 25.02.17.

Assenmacher, D.; Clever, L.; Frischlich, L.; Quandt, T.; Trautmann, H.; Grimme, C. (2020). *Demystifying Social Bots: On the Intelligence of Automated Social Media Actors.* Social Media + Society 6 (3), S.1-14. SAGE Publications, Thousand Oaks.

Blank, H.; Launay, C. (2014). *How to protect eyewitness memory against the misinformation effect: A meta-analysis of post-warning studies*. Journal of Applied Research in Memory and Cognition 3 (2), S.77-88. Elsevier, Amsterdam.

Blom, J.; Hansen, K. (2015). *Click bait: Forward-reference as lure in online news headlines.* Journal of Pragmatics 76, S.87-100. Elsevier, Amsterdam.

Boehm, L. (1994). *The Validity Effect: A Search for Mediating Variables*. Personality and Social Psychology Bulletin 20 (3), S.285-293. SAGE Publications, Thousand Oaks.

Borchers, C. (2017). *"Fake News" Has Now Lost All Meaning*. Online im Internet: https://www.washingtonpost.com/news/the-fix/wp/2017/02/09/fake-news-has-now-lost-all-meaning/?utm_term=.70fa3df16f17 (Abrufdatum: 24.08.2020), veröffentlicht am 09.02.2017.

Boshmaf, Y.; Muslukhov, I.; Beznosov, K. Ripeanu, M. (2013). *Design and Analysis of a Social Botnet*. Computer Networks 57 (2), S.556–578. Elsevier, Amsterdam.

Bounegru, L.; Gray, J.; Venturini, T.; Mauri, M. (2018*): A Field Guide To "Fake News" And Other Information Disorders*. Online im Internet: https://fakenews.publicdatalab.org (Abrufdatum: 13.09.2020), veröffentlicht im April 2019. Public Data Lab, Amsterdam.

Boutyline, A.; Willer, R. (2017). *The Social Structure of Political Echo Chambers: Variation in Ideological Homophily in Online Networks.* Political Psychology 38 (3), S.551-569. Wiley, Hoboken.

Brachten, F.; Mirbabaie, M.; Stieglitz, S.; Berger, O.; Bludau, S.; Schrickel, K. (2018). *Threat or Opportunity? - Examining Social Bots in Social Media Crisis Communication*. Online im Internet: http://arxiv.org/pdf/1810.09159v1 (Abrufdatum: 24.09.20), veröffentlicht am 22.10.18.

Bronstein, M.; Pennycook, G.; Bear, A.; Rand, D.; Cannon, T. (2019). *Belief in Fake News is Associated with Delusionality, Dogmatism, Religious Fundamentalism, and Reduced Analytic Thinking.* Journal of Applied Research in Memory and Cognition 8 (1), S.108-117. Elsevier, Amsterdam.

Burgoon, J.; Blair, J.; Qin, T.; Nunamaker, J. (2003). *Detecting Deception through Linguistic Analysis*. In: Goos, G.; Hartmanis, J.; van Leeuwen, J.; Chen, H.; Miranda, R.; Zeng, D.; Demchak, C.; Schroeder, J.; Madhusudan, T. (Hg.): Intelligence and Security Informatics, S.91-101. Springer, Berlin.

Burkart, R. (2019). *Kommunikationswissenschaft: Grundlagen und Problemfelder einer interdisziplinären Sozialwissenschaft.* 5. Auflage. Böhlau Verlag, Wien, Köln, Weimar.

Cambria, E.; Schuller, B.; Xia, Y.; Havasi, C. (2013). *New Avenues in Opinion Mining and Sentiment Analysis.* IEEE Intelligent Systems 28 (2), S.15-21. IEEE, Piscataway Township.

Castillo, C.; Mendoza, M.; Poblete, B. (2011). *Information credibility on twitter*. In: Sadagopan, S.; Ramamritham, K.; Kumar, A.; Ravindra, M.; Bertino, E.; Kumar, R. (Hg.): Proceedings of the 20th international conference on World Wide Web, S.675-684. ACM Press, New York.

Chhabra, S.; Aggarwal, A.; Benevenuto, F.; Kumaraguru, P. (2011). *Phi.sh/$oCiaL: The phishing landscape through short URLs.* In: Potdar, V. (Hg.): Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference on - CEAS '11, S.92-101. ACM Press, New York.

Chu, Z.; Gianvecchio, S.; Wang, H.; Jajodia, S. (2012*). Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?* IEEE Transactions on Dependable and Secure Computing 9 (6), S.811–824. IEEE, Piscataway Township.

Ciampaglia, G.; Shiralkar, P.; Rocha, L.; Bollen, J.; Menczer, F.; Flammini, A. (2015). *Computational Fact Checking from Knowledge Networks.* PloS one 10 (6). PLOS, San Francisco.

Cobb, M.; Nyhan, B.; Reifler, J. (2013). *Beliefs Don't Always Persevere: How Political Figures Are Punished When Positive Information about Them Is Discredited*. Political Psychology 34 (3), S.307-326. Wiley, Hoboken.

Colleoni, E.; Rozza, A.; Arvidsson, A. (2014). *Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data.* Journal of Communication 64 (2), S.317-332. Wiley, Hoboken.

Conroy, N.; Rubin, V.; Chen, Y. (2015). *Automatic Deception Detection: Methods for Finding Fake News*. Proceedings of the Association for Information Science and Technology 52 (1), S.1-4. Wiley, Hoboken.

Davidson, W. (1983). *The Third-Person Effect in Communication*. The Public Opinion Quarterly 47 (1), S.1-15. Oxford University Press, Oxford.

Davis, C.; Varol, O.; Ferrara, E.; Flammini, A.; Menczer, F. (2016). *BotOrNot: A System to Evaluate Social Bots*. Proceedings of the 25th International Conference Companion on World Wide Web, S.273-274. ACM Press, New York.

Della Vedova, M.; Tacchini, E.; Moret, S.; Ballarin, G.; DiPierro, M.; Alfaro, L. (2018). *Automatic Online Fake News Detection Combining Content and Social Signals*. 22nd Conference of Open Innovations Association, S.272–279. IEEE, Piscataway Township.

Fallis, D. (2015). *What Is Disinformation?* Library Trends 63 (3), S.401-426. JHU Press, Baltimore.

Fazio, L.; Rand, D.; Pennycook, G. (2019). *Repetition increases perceived truth equally for plausible and implausible statements.* Psychonomic Bulletin & Review 26 (5), S.1705-1710. Springer, Berlin.

Feng, S.; Banerjee, R.; Choi, Y. (2012). *Syntactic stylometry for deception detection*. In: Li, H.; Lin, C.; Osborne, M.; Lee, G.; Park, J. (Hg.): Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), S.171-175. Association for Computational Linguistics, Stroudsburg.

Ferrara, E.; Varol, O.; Davis, C.; Menczer, F.; Flammini, A. (2016). *The Rise of Social Bots*. Communications of the ACM 59 (7), S.96–104. ACM Press, New York.

Fiskkit (2020). *What is "Fisking"?* Online im Internet: https://fiskkit.com/about (Abrufdatum: 15.10.20).

Fraunhofer FKIE (2019). *Software für die automatisierte Erkennung von Fake News*. Online im Internet: https://www.fraunhofer.de/de/presse/presseinformationen/2019/februar/software-fuer-die-automatisierte-erkennung-von-fake-news.html (Abrufdatum: 24.09.20), veröffentlicht am 01.02.19.

Freeze, M.; Baumgartner, M.; Bruno, P.; Gunderson, J.; Olin, J.; Ross, M.; Szafran, J. (2020). *Fake Claims of Fake News: Political Misinformation, Warnings, and the Tainted Truth Effect.* Political Behaviour 2020. Springer, Berlin.

Freitas, C.; Benevenuto, F.; Veloso, A.; Ghosh, S. (2016). *An empirical study of socialbot infiltration strategies in the Twitter social network.* Social Network Analysis and Mining 6 (1). Springer, Berlin.

Full Fact (2020). *Automated Fact Checking*. Online im Internet: https://full fact.org/about/automated/ (Abrufdatum: 27.10.20).

Golbeck, J.; Robles, C.; Edmondson, M.; Turner, K. (2011). *Predicting Personality from Twitter*. 2011 IEEE third international conference on Privacy, Security, Risk and Trust and 2011 IEEE third international conference on Social Computing, S.149–156. IEEE, Piscataway Township.

Graves, L. (2018). *Understanding the Promise and Limits of Automated Fact-Checking*. Online im Internet: https://reutersinstitute.politics.ox.ac.uk/our-research/understanding-promise-and-limits-automated-fact-checking (Abrufdatum: 05.10.20), veröffentlicht am 26.02.18.

Guess, A.; Nagler, J.; Tucker, J. (2019). *Less than you think: Prevalence and predictors of fake news dissemination on Facebook.* Science Advances 5 (1). American Association for the Advancement of Science, Washington D.C.

Gupta, A.; Kumaraguru, P.; Castillo, C.; Meier, P. (2014). *TweetCred: Real-Time Credibility Assessment of Content on Twitter*. In: Aiello, L.; McFarland, D. (Hg.): Social Informatics 8851, S.228-243. Springer, Berlin.

Gupta, A.; Lamba, H.; Kumaraguru, P.; Joshi, A. (2013). *Faking Sandy: characterizing and identifying fake images on Twitter during Hurricane Sandy.* In: Schwabe, D.; Almeida, V.; Glaser, H.; Baeza-Yates, R.; Moon, S. (Hg.): Proceedings of the 22nd International Conference on World Wide Web - WWW '13 Companion, S.729–736. ACM Press, New York.

Gupta, M.; Zhao, P.; Han, J. (2012). *Evaluating Event Credibility on Twitter*. In: Ghosh, J.; Liu, H.; Davidson, I.; Domeniconi, C.; Kamath, C. (Hg.): Proceedings of the 2012 SIAM International Conference on Data Mining, S.153–164. Society for Industrial and Applied Mathematics, Philadelphia.

Hanselowski, A.; PVS, A.; Schiller, B.; Caspelherr, F.; Chaudhuri, D.; Meyer, C.; Gurevych, I. (2018). *A Retrospective Analysis of the Fake News Challenge Stance Detection Task.* Online im Internet: http://arxiv.org/pdf/1806.05180v1 (Abrufdatum: 20.10.20), veröffentlicht am 13.06.18.

Hassan, N.; Adair, B.; Hamilton, J.; Li, C.; Tremayne, M.; Yang, J.; Yu, C. (2015). *The Quest to Automate Fact-Checking*. Proceedings of the 2015 Computation + Journalism Symposium. Evanston.

Hassan, N.; Zhang, G.; Arslan, F.; Caraballo, J.; Jimenez, D.; Gawsane, S.; Hasan, S.; Joseph, M.; Kulkarni, A.; Nayak, A.; Sable, V.; Li, C.; Tremayne, M. (2017*). ClaimBuster: The First-ever End-to-end Fact-checking System*. Proceedings of the VLDB Endowment 10 (12), S.1945–1948. VLDB Endowment, Stanford.

Hegelich, S. (2016). *Invasion der Meinungs-Roboter.* Analysen & Argumente 221. Konrad-Adenauer-Stiftung, Sankt Augustin.

Hegelich, S.; Janetzko, D. (2016). *Are Social Bots on Twitter Political Actors? Empirical Evidence from a Ukrainian Social Botnet*. In: Strohmaier, M.; Gummadi, K.; Lindner, D.; Weller, K.; Gilbert, E.; Macy, M.; Wagner, C. (Hg.): Proceedings of the Tenth International AAAI Conference on Web and Social Media, S. 579–582. Association for the Advancement of Artificial Intelligence, Menlo Park.

Hen, Jacqueline (2020). *Automatisierte Wahrheitssuche - Software soll Fake News erkennen.* Online im Internet: https://www.fraunhofer-innovisions.de/semantische-medienanalyse/automatisierte-wahrheitssuche/ (Abrufdatum: 20.10.20), veröffentlicht am 20.02.20.

Hernon, P. (1995). *Disinformation and misinformation through the internet: Findings of an exploratory study*. Government Information Quarterly 12(2), 133–139. Elsevier, Amsterdam.

Holan, A. (2018). *The Principles of the Truth-O-Meter: PolitiFact's methodology for independent fact-checking.* Online im Internet: https://www.politifact.com/article/2018/feb/12/principles-truth-o-meter-politifacts-methodology-i/ (Abrufdatum: 15.10.20), veröffentlicht am 12.02.18.

Horne, B.; Adali, S. (2017). *This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News*. Online im Internet: https://arxiv.org/abs/1703.09398 (Abrufdatum: 13.09.2020), veröffentlicht am 28.03.2017.

Howard, P.; Woolley, S.; Calo, R. (2018). *Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration.* Journal of Information Technology & Politics 15 (2), S.81-93. American Political Science Association, Washington D.C.

Huiwen, N. (2018). *Most people say they can spot fake news but falter when tested: Survey*. Online im Internet: https://www.straitstimes.com/singapore/most-people-say-they-can-spot-fake-news-but-falter-when-tested-survey (Abrufdatum: 22.09.20), veröffentlicht am 28.09.18.

Islam, M.; Sarkar, T.; Khan, S.; Mostofa Kamal, A.; Hasan, S.; Kabir, A.; Yeasmin, D.; Islam, M. A.; Amin Chowdhury, K.; Anwar, K.; Chughtai, A.; Seale, H. (2020).*COVID-19-Related Infodemic and Its Impact on Public Health: A Global Social Media Analysis*. The American Journal of Tropical Medicine and Hygiene 00(0) 2020, S.1-9. The American Society of Tropical Medicine and Hygiene, Illinois.

Isola, P.; Zhu, J.; Zhou, T.; Efros, A. (2016). *Image-to-Image Translation with Conditional Adversarial Networks*. Online im Internet: http://arxiv.org/pdf/1611.07004v3 (Abrufdatum: 28.09.20), veröffentlicht am 21.11.16.

Jack, C. (2017). *Lexicon of Lies: Terms for Problematic Information*. Online im Internet: https://datasociety.net/pubs/oh/DataAndSociety_LexiconofLies.pdf (Abrufdatum: 24.08.2020), veröffentlicht am 09.08.2017.

Jin, Z.; Cao, J.; Zhang, Y.; Zhou, J.; Lou, J. (2016). *News Verification by Exploiting Conflicting Social Viewpoints in Microblogs*. In: Schuurmans, D.; Wellman, M. (Hg.): Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, S.2972-2978. Association for the Advancement of Artificial Intelligence, Menlo Park.

Jin, Z.; Cao, J.; Zhang, Y.; Zhou, J.; Tian, Q. (2017). *Novel Visual and Statistical Image Features for Microblogs News Verification*. IEEE Transactions on Multimedia 19 (3), S.598–608. IEEE, Piscataway Township.

Johnson, H.; Seifert, C. (1994). *Sources of the Continued Influence Effect: When Misinformation in Memory Affects Later Inferences.* Journal of Experimental Psychology: Learning, Memory, and Cognition 20 (6), S.1420-1436. American Psychological Association, Washington D.C.

Kalchbrenner, N.; Elsen, E.; Simonyan, K.; Noury, S.; Casagrande, N.; Lockhart, E.; Stimberg, F.; van den Oord, A.; Dieleman, S.; Kavukcuoglu, K. (2018). Efficient *Neural Audio Synthesis.* Online im Internet: https://arxiv.org/pdf/1802.08435.pdf (Abrufdatum: 29.09.20), veröffentlicht am 25.06.18.

Kapusta, J.; Obonya, J. (2020). *Improvement of Misleading and Fake News Classification for Flective Languages by Morphological Group Analysis.* Informatics 7 (1), Nr.4. MDPI, Basel.

Kietzmann, J.; Lee, L.; McCarthy, I.; Kietzmann, T. (2020). *Deepfakes: Trick or treat?* Business Horizons 63 (2), S.135-146. Elsevier, Amsterdam.

Kind, S.; Bovenschulte, M.; Ehrenberg-Silies, S.; Jetzke, T.; Weide, S. (2017). *Social Bots.* Online im Internet: https://www.tab-beim-bundestag.de/de/aktuelles/20161219/Social%20Bots_Thesenpapier.pdf. (Abrufdatum: 23.09.20), veröffentlicht im Januar 2017. Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag, Berlin.

Klein, D.; Wueller, J. (2017): *Fake News: A Legal Perspective*. Journal of Internet Law 20 (10), S.1,6-13, Aspen Publishers, New York.

Knight, W. (2019). *An AI that writes convincing prose risks mass-producing fake news*. Online im Internet: https://www.technologyreview.com/2019/02/14/137426/an-ai-tool-auto-generates-fake-news-bogus-tweets-and-plenty-of-gibberish/ (Abrufdatum: 28.09.20), veröffentlicht am 14.02.19.

Kosinski, M.; Stillwell, D.; Graepel, T. (2013). *Private traits and attributes are predictable from digital records of human behavior*. Proceedings of the National Academy of Sciences of the United States of America 110 (15), S.5802-5805. National Academy of Sciences, Washington D.C.

Krosnick, J.; Alwin, D. (1987). *An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement*. Public Opinion Quarterly 51 (2), S.201–219. Oxford University Press, Oxford.

Kuran, T.; Sunstein, C. (1999). *Availability Cascades and Risk Regulation*. Stanford Law Review 51 (4), S.683–768. Stanford Law School, Stanford.

Laaff, M. (2019). *Hello, Adele – bist du's wirklich?* Online im Internet: https://www.zeit.de/digital/internet/2019-11/deepfakes-gefaelschte-videos-kuenstliche-intelligenz-manipulation (Abrufdatum: 29.09.20), veröffentlicht am 10.11.19.

Lazer, D.; Baum, M.; Benkler, Y.; Berinsky, A.; Greenhill, K.; Menczer, F.; Metzger, M.; Nyhan, B.; Pennycook, G.; Rothschild, D.; Schudson, M.; Sloman, S.; Sunstein, C.; Thorson, E.; Watts, D.; Zittrain, J. (2018): *The science of fake news*. Science 359 (6380), S. 1094–1096. American Association for the Advancement of Science, Washington D.C.

Lazer, D.; Baum, M.; Grinberg, N.; Friedland, L.; Joseph, K.; Hobbs, W.; Mattsson, C. (2017). *Combating fake news: An agenda for research and action.* Online im Internet: https://shorensteincenter.org/wp-content/uploads/2017/05/Combating-Fake-News-Agenda-for-Research-1.pdf (Abrufdatum: 29.09.20), veröffentlicht am 02.05.17.

Luber, S.; Litzel, N. (2018). *Was ist ein Chatbot?* Online im Internet: https://www.bigdata-insider.de/was-ist-ein-chatbot-a-690591/ (Abrufdatum: 28.09.20), veröffentlicht am 28.02.18.

Marchal, N.; Kollanyi, B; Neudert, L.; Howard, P. (2019). *Junk News During the EU Parliamentary Elections: Lessons from a Seven-Language Study of Twitter and Facebook*. Online im Internet: https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/05/EU-Data-Memo.pdf (Abrufdatum: 22.09.20), veröffentlicht am 20.05.19. University of Oxford, Oxford.

McManus, C.; Michaud, C. (2018). *Never Mind the Buzzwords: Defining Fake News and Post-Truth*. King's Centre for Strategic Communications/NATO Strategic Communications Centre of Excellence (Hg.), Fake News - A Roadmap, S.14–20. King's Centre for Strategic Communications; NATO Strategic Communications Centre of Excellence, Riga, London.

Metzger, M.; Flanagin, A.; Medders, R. (2010*). Social and Heuristic Approaches to Credibility Evaluation Online.* Journal of Communication 60 (3), S.413-439. Wiley, Hoboken.

Morris, M.; Counts, S.; Roseway, A.; Hoff, A.; Schwarz, J. (2012). *Tweeting is Believing?: Understanding Microblog Credibility Perceptions* In: Poltrock, S.; Simone, C.; Grudin, J.; Mark, G.; Riedl, J. (Hg.): Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12, S.441-450. ACM Press, New York.

Mullen, B.; Brown, R.; Smith, C. (1992). *Ingroup bias as a function of salience, relevance, and status: An integration.* European Journal of Social Psychology 22 (2), S.103-122. Wiley, Hoboken.

Newman, M.; Pennebaker, J.; Berry, D.; Richards, J. (2003). *Lying words: predicting deception from linguistic styles.* Personality and Social Psychology Bulletin 29 (5), S.665–675. SAGE Publications, Thousand Oaks.

Newman, N.; Fletcher, R.; Schulz, A.; Andi, S. Nielsen, R. (2020). *Digital News Report 2020*. Online im Internet: http://www.digitalnewsreport.org/ (Abrufdatum: 21.09.20), veröffentlicht am 15.07.20. Reuters Institut, Oxford.

NewsGuard (2020). *Bewertungsprozess und Kriterien*. Online im Internet: https://www.newsguardtech.com/de/bewertungen/bewertungsprozess-und-kriterien/ (Abrufdatum: 20.10.20).

Nickerson, R. (1998). *Confirmation Bias: A Ubiquitous Phenomenon in Many Guises*. Review of General Psychology 2 (2), S.175–220. Eductional Pu blishing Foundation, Washington D.C.

Nyhan, B.; Reifler, J. (2010). *When Corrections Fail: The Persistence of Political Misperceptions*. Political Behaviour 32 (2), S.303-330. Springer, Berlin.

Office of Cyber and Infrastructure Analysis (2018). *Social Media Bots Overview.* Online im Internet: https://www.cisa.gov/sites/default/files/publications/19_0717_cisa_social-media-bots-overview.pdf (Abrufdatum: 23.09.20), veröffentlicht im Mai 2018. Homeland Security, Washington D.C.

Pariser, E. (2012). *Filter Bubble: Wie wir im Internet entmündigt werden.* Hanser, München.

Paskin, D. (2018): Real or Fake News: Who Knows? The Journal of Social Media in Society, 7 (2), S.252-273, Tarleton State University, Stephenville.

Pennycook, G.; Bear, A.; Collins, E.; Rand, D. (2020). *The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings.* Management Science, Articles in Advance, S.1-14. Informs, Catonsville.

Pennycook, G.; Rand, D. (2019). *Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning*. Cognition 188, S.39-50. Elsevier, Amsterdam.

Pennycook, G.; Rand, D. (2020). *Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking.* Journal of Personality 88 (2), S.185-200. Wiley, Hoboken.

Pérez-Rosas, V.; Kleinberg, B.; Lefevre, A.; Mihalcea, R. (2018). *Automatic Detection of Fake News.* Proceedings of the 27th International Conference on Computational Linguistics, S.3391-3401. Association for Computational Linguistics, Stroudsburg.

Perov, I.; Gao, D.; Chervoniy, N.; Liu, K.; Marangonda, S.; Umé, C.; Dpfks; Facenheim, C.; RP, L.; Jiang, J.; Zhang, S.; Wu, P.; Zhou, B.; Zhang, W. (2020). *DeepFaceLab: A simple, flexible and extensible face swapping framework.* Online im Internet: http://arxiv.org/pdf/2005.05535v4 (Abrufdatum: 29.09.20), veröffentlicht am 12.05.20.

Popat, K.; Mukherjee, S.; Yates, A.; Weikum, G. (2018). *DeClarE: Debunking Fake News and False Claims using Evidence-Aware Deep Learning.* Online im Internet: http://arxiv.org/pdf/1809.06416v1 (Abrufdatum: 20.10.20), veröffentlicht am 17.09.18.

Potthast, M.; Kiesel, J.; Reinartz, K.; Bevendorff, J.; Stein, B. (2018). *A Stylometric Inquiry into Hyperpartisan and Fake News.* In: Gurevych, I.; Miyao, Y. (Hg.): Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), S.231–240. Association for Computational Linguistics, Stroudsburg.

PricewaterhouseCoopers (2018). *Vertrauen in Medien.* Online im Internet: https://www.pwc.de/de/technologie-medien-und-telekommunikation/pwc-studie-vertrauen-in-medien-2018.pdf (Abrufdatum: 21.09.2020), veröffentlicht im Juli 2018.

PricewaterhouseCoopers (2019). *„Fake News" - Ergebnisse einer Bevölkerungsbefragung*. Online im Internet: https://www.pwc.de/de/technologie-medien-und-telekommunikation/pwc-bevoelkerungsbefragung-fake-news.pdf (Abrufdatum: 10.09.2020), veröffentlicht im April 2019.

Reece, A.; Danforth, C. (2017). *Instagram photos reveal predictive markers of depression*. EPJ Data Science 6 (1). Springer, Berlin.

Reis, J.; Correia, A.; Murai, F.; Veloso, A.; Benevenuto, F.; Cambria, E. (2019). *Supervised Learning for Fake News Detection.* IEEE Intelligent Systems 34 (2), S.76-81. IEEE Computer Society, Washington D.C.

Roberts, E. (2020). *Bad Bot Report 2020: Bad Bots Strike Back.* Online im Internet: https://www.imperva.com/blog/bad-bot-report-2020-bad-bots-strike-back/ (Abrufdatum: 23.09.2020), veröffentlicht am 21.04.20.

Ross, L.; Lepper, M.; Hubbard, M. (1975.) *Perseverance in Self-Perception and Social Perception: Biased Attributional Processes in the Debriefing Paradigm*. Journal of Personality and Social Psychology 32(5), S.880-892. American Psychological Association, Washington D.C.

Roth, Y.; Pickles, N. (2020). *Updating our approach to misleading information*. Online im Internet: https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html (Abrufdatum: 21.10.20), veröffentlicht am 11.05.20.

Ruchansky, N.; Seo, S.; Liu, Y. (2017). *CSI - A Hybrid Deep Model for Fake News Detection*. In: Lim, E.; Winslett, M.; Sanderson, M.; Fu, A.; Sun, J.; Culpepper, S.; Lo, E.; Ho, J.; Donato, D.; Agrawal, R.; Zheng, Y.; Castillo, C.; Sun, A.; Tseng, V.; Li, C. (Hg.): Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, S.797–806. ACM Press, New York.

Ruddick, G. (2017*). Experts sound alarm over news websites' fake news twins*. Online im Internet: https://www.theguardian.com/technology/2017/aug/18/experts-sound-alarm-over-news-websites-fake-news-twins (Abrufdatum: 29.09.20), veröffent licht am 18.08.17.

Schäffer, B. (2007). *The Digital Literacy of Seniors.* Research in Comparative and Inter national Education 2 (1), S.29-42. SAGE Publications, Thousand Oaks.

Seymour, J.; Tully, P. (2016). *Weaponizing Data Science for Social Engineering: Automated E2E Spear Phishing on Twitter*. Online im Internet: https://www.black hat.com/docs/us-16/materials/us-16-Seymour-Tully-Weaponizing-Data-Science-For-Social-Engineering-Automated-E2E-Spear-Phishing-On-Twitter-wp.pdf (Abrufdatum: 24.09.20), veröffentlicht im Jahr 2016. DEF CON, Las Vegas.

Shariff, S.; Zhang, X.; Sanderson, M. (2017). *On the credibility perception of news on Twitter: Readers, topics and features.* Computers in Human Behavior 75, S. 785–796. Elsevier, Amsterdam.

Shi, B.; Weninger, T. (2016). *Fact Checking in Heterogeneous Information Networks*. In: Bourdeau, J.; Hendler, J..; Nkambou, R.; Horrocks, I.; Zhao, B. (Hg.): Proceedings of the 25th International Conference Companion on World Wide Web - WWW '16 Companion, S.101-102. ACM Press, New York.

Shu, K.; Sliva, A.; Wang, S.; Tang, J.; Liu, H. (2017). *Fake News Detection on Social Media: A Data Mining Perspective.* Online im Internet: http://ar xiv.org/pdf/1708.01967v3 (Abrufdatum: 13.10.20), veröffentlicht am 07.08.17.

Snopes (2020). *Snopes.com follows all industry guidelines for transparency in reporting.* Online im Internet: https://www.snopes.com/transparency/ (Abrufdatum: 15.10.20).

Sparrow, B.; Liu, J.; Wegner, D. (2011). *Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips*. Science 333 (6043), S.776-778. American Association for the Advancement of Science, Washington D.C.

Statistisches Bundesamt (2020). *Personen mit Internetaktivitäten zu privaten Zwecken nach Alter*. Online im Internet: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Einkommen-Konsum-Lebensbedingungen/IT-Nutzung/Tabellen/internetaktivitaeten-personen-alter-ikt.html (Abrufdatum: 22.09.20), veröffentlicht am 11.08.20.

Stefanowitsch, A. (2017). *Anglizismus des Jahres 2016*. Online im Internet: http://www.anglizismusdesjahres.de/anglizismen-des-jahres/adj-2016/ (Abrufdatum: 10.09.2020).

Stieglitz, S.; Brachten, F.; Berthelé, D.; Schlaus, M.; Venetopoulou, C.; Veutgen, D. (2017). *Do Social Bots (Still) Act Different to Humans? – Comparing Metrics of Social Bots with Those of Humans*. In: Meiselwitz, G. (Hg.): Social Computing and Social Media. Human Behavior 10282, S. 379–395. Springer, Berlin.

Stieglitz, S.; Brachten, F.; Ross, B.; Jung, A. (2017*). Do Social Bots Dream of Electric Sheep? A Categorisation of Social Media Bot Accounts*. ACIS 2017 Proceedings. 89. Association for Information Systems, Atlanta.

Stocké, V. (2002). *Framing und Rationalität. die Bedeutung der Informationsdarstellung für das Entscheidungsverhalten*. Oldenbourg, München.

Subrahmanian, V.; Azaria, A.; Durst, S.; Kagan, V.; Galstyan, A.; Lerman, K.; Zhu, L.; Ferrara, E.; Flammini, A.; Menczer, F. (2016). *The DARPA Twitter Bot Challenge*. Computer 49 (6), S.38–46. IEEE, Piscataway Township.

Sullivan, M. (2017). *It's time to retire the tainted term 'fake news'*. Online im Internet: https://www.washingtonpost.com/lifestyle/style/its-time-to-retire-the-tainted-term-fake-news/2017/01/06/a5a7516c-d375-11e6-945a-76f69a399dd5_story.html (Abrufdatum: 13.09.2020), veröffentlich am 08.01.2017.

Sundar, S.; Oeldorf-Hirsch, A.; Xu, Q. (2008). *The bandwagon effect of collaborative filtering technology.* In: Czerwinski, Lund et al. (Hg.) 2008 – Proceeding of the twenty-sixth annual CHI conference extended abstracts on Human factors in computing systems - CHI '08, S. 3453–3458. 05.04.2008 - 10.04.2008. Florence, Italien. ACM Press, New York.

Swire, B.; Berinsky, A.; Lewandowsky, S.; Ecker, U. (2017). *Processing political misinformation: comprehending the Trump phenomenon.* Royal Society open science 4 (3). Royal Society Publishing, London.

Tacchini, E.; Ballarin, G.; Della Vedova, M.; Moret, S.; Alfaro, L. (2017). *Some Like it Hoax: Automated Fake News Detection in Social Networks.* Online im Internet: http://arxiv.org/pdf/1704.07506v1 (Abrufdatum: 21.10.2020), veröffentlicht am 25.04.17.

Tandoc, E.; Lim, Z.; Ling, R. (2018). *Defining "Fake News".* Digital Journalism 6 (2), S.137–153, Taylor & Francis, London.

Tversky, A.; Kahneman, D. (1973). *Availability: A heuristic for judging frequency and probability.* Cognitive Psychology 5 (2), S.207-232. Elsevier, Amsterdam.

UK Parliament (2018*). Disinformation and 'fake news': Interim Report.* Online im Internet: https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/363/36304.htm (Abrufdatum: 24.08.2020), veröffentlicht am 29.07.2018.

Vallone, R.; Ross, L.; Lepper, M. (1985). *The Hostile Media Phenomenon: Biased Perception and Perceptions of Media Bias in Coverage of the Beirut Massacre*. Journal of Personality and Social Psychology 49 (3), S.577–585. American Psychological Association, Washington D.C.

Van Damme, I.; Smets, K. (2014). *The power of emotion versus the power of suggestion: memory for emotional events in the misinformation paradigm*. Emotion 14 (2), S.310-320. American Psychological Association, Washington D.C.

Vosoughi, S.; Roy, D.; Aral, S. (2018). *The spread of true and false news online*. Science 359 (6380), S.1146-1151. American Association fort he Advancement of Science, Washington D.C.

Voss, K. (2010): *Grassrootscampaigning und Chancen durch neue Medien.* Online im Internet: https://www.bpb.de/apuz/32777/grassrootscampaigning-und-chancen-durch-neue-medien (Abrudfatum: 24.09.20), veröffentlicht am 03.05.10.

Wang, W. (2017). *„Liar, Liar Pants on Fire“: A New Benchmark Dataset for Fake News Detection.* Online im Internet: http://arxiv.org/pdf/1705.00648v1 (Abrufdatum: 08.10.20), veröffentlicht am 01.05.17.

Ward, E. (1982). *Conservatism in Human Information Processing*. In: Kahneman, D.; Slovic, P.; Tversky,A. (Hg.): Judgment under uncertainty: Heuristics and biases, S.359-369. Cambridge University Press, Cambridge.

Wardle, C. (2017). *Fake News. It's complicated.* Online im Internet: https://first draftnews.org/latest/fake-news-complicated/ (Abrufdatum: 10.09.2020), veröffentlicht am 16.02.2017.

Wardle, C.; Derakhshan, H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making.* Council of Europe report DGI(2017)09. Online im Internet: https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c (Abrufdatum: 24.08.2020), veröffentlicht am 27.09.2017.

Weedon, J.; Nuland, W.; Stamos, A. (2017). *Information Operations and Facebook*. Online im Internet: https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf (Abrufdatum: 21.10.20), veröffentlicht am 27.04.17.

Weeks, B. (2015). *Emotions, Partisanship, and Misperceptions: How Anger and Anxiety Moderate the Effect of Partisan Bias on Susceptibility to Political Misinformation*. Journal of Communication 65 (4), S.699-719. Wiley, Hoboken.

Wei, X.; Stillwell, D. (2017). *How smart does your profile image look? Estimating intelligence from social network profile images.* Proceedings of the tenth ACM international conference on web search and data mining, S. 33–40. ACM Press, New York.

Wineburg, S.; McGrew, S.; Breakstone, J.; Ortega, T. (2016). *Evaluating Information: The Cornerstone of Civic Online Reasoning*. Online im Internet: https://purl.stanford.edu/fv751yt5934 (Abrufdatum: 22.09.20), veröffentlich am 22.10.16. Stanford University, Stanford.

Wojciszke, B.; Brycz, H.; Borkenau, P. (1993). *Effects of Information Content and Evaluative Extremity on Positivity and Negativity Biases.* Journal of Personality and Social Psychology 64 (3), S.327-335. American Psychological Association, Washington D.C.

Wu, Y.; Agarwal, P.; Li, C.; Yang, J.; Yu, C. (2014). *Toward computational fact-che cking.* Proceedings of the VLDB Endowment 7 (7), S.589-600. VLDB Endowment, Stanford.

Wu, Y.; Walenz, B.; Li, P.; Shim, A.; Sonmez, E.; Agarwal, P.; Li, C.; Yang, J.; Yu, C. (2014). *iCheck: Computationally Combating "Lies, D—ned Lies, and Statistics".* In: Dyreson, C.; Li, F.; Özsu, M. (Hg.): Proceedings of the 2014 ACM SIGMOD international conference on Management of data, S.1063–1066. ACM Press, New York.

Yang, J.; Counts, S.; Morris, M.; Hoff, A. (2013). *Microblog Credibility Perceptions: Comparing the United States and China.* In: Bruckman, A.; Counts, S.; Lampe, C.; Terveen, L. (Hg.): Proceedings of the 2013 conference on Computer supported cooperative work - CSCW '13, S.575-586. ACM Press, New York.

Yang, S.; Shu, K.; Wang, S.; Gu, R.; Wu, F.; Liu, H. (2019). *Unsupervised Fake News Detection on Social Media: A Generative Approach.* Proceedings of the AAAI Conference on Artificial Intelligence 33, S.5644–5651. Association for the Advancement of Artificial Intelligence, Menlo Park.

Yang, Y.; Zheng, L.; Zhang, J.; Cui, Q.; Li, Z.; Yu, P. (2018). *TI-CNN: Convolutional Neural Networks for Fake News Detection.* Online im Internet: http://arxiv.org/pdf/1806.00749v1 (Abrufdatum: 22.09.20), veröffentlicht am 03.06.18.

YouGov (2017). *Alles Fake?!* Online im Internet: https://campaign.yougov.com/DE_2017_08_Political_Fake_News.html (Abrufdatum: 21.09.20), veröffentlicht im August 2017.

Zellers, R.; Holtzman, A.; Rashkin, H.; Bisk, Y.; Farhadi, A.; Roesner, F.; Choi, Y. (2019). *Defending Against Neural Fake News*. Online im Internet: http://arxiv.org/pdf/1905.12616v2 (Abrufdatum: 29.09.20), veröffentlicht am 29.05.19.

Zhang, H.; Xu, T.; Li, H.; Zhang, S.; Wang, X.; Huang, X.; Metaxas, D. (2017). *Stack-GAN: Text to Photo-Realistic Image Synthesis with Stacked Generative Adversarial Networks.* 2017 IEEE International Conference on Computer Vision, S.5908–5916. IEEE, Piscataway Township.

Zhou, L.; Burgoon, J.; Nunamaker, J.; Twitchell, D. (2004). *Automating Linguistics-Based Cues for Detecting Deception in Text-Based Asynchronous Computer-Mediated Communications.* Group Decision and Negotiation 13 (1), S.81-106. Springer, Berlin.

Zubiaga, A.; Liakata, M.; Procter, R.; Wong Sak Hoi, G.; Tolmie, P. (2016). *Analysing How People Orient to and Spread Rumours in Social Media by Looking at Conversational Threads.* PLoS one 11(3). Public Library of Science, San Francisco.

Zuiderveen Borgesius, F.; Trilling, D.; Möller, J.; Bodó, B.; Vreese, C.; Helberger, N. (2016). *Should We Worry about Filter Bubbles?* Internet Policy Review 5 (1), S.1–16. Humboldt Institut für Internet und Gesellschaft, Berlin.

Eidesstattliche Erklärung

Ich erkläre hiermit eidesstattlich, dass ich die vorliegende Arbeit selbständig angefertigt habe. Die aus fremden Quellen übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht. Ich bin mir bewusst, dass eine unwahre Erklärung rechtliche Folgen haben kann.

_____ _____

Ort, Datum Vorname und Nachname

Veröffentlichung der Arbeit

Ich stimme der Veröffentlichung der Arbeit im Rahmen von Forschung und Lehre an der Friedrich-Schiller-Universität Jena zu.

_____ _____

Ort, Datum Vorname und Nachname